CHAPTER 6
BIG BANG COSMOLOGY – THE EVOLVING UNIVERSE

6.1 Introduction

In Chapter 5 we saw how general relativity can be used to construct models of the Universe. These models describe how the scale factor varies in a Universe that is filled with a smooth distribution of matter and radiation, but say very little about the properties and behaviour of these components. So, for instance, they do not account for the consequences of microscopic processes such as the interactions between the particles that make up the matter within the Universe.

However, such interactions play an important role in cosmology. One example that we shall see later in this chapter is that cosmological theories can offer an explanation for the observation that most of the stars in the Universe have a composition that is approximately 75% hydrogen and 25% helium (by mass). A fundamental aspect of the process that forms helium is that it involves reactions between nuclei and particles at an early stage in the history of the Universe. Thus physical processes on small scales *must* be taken into consideration in order to develop an understanding of the evolution of the Universe. This chapter starts by examining the conditions under which particles interacted in the early Universe (Section 6.2). The main focus, however, is to follow a chronological sequence from very early times in the history of the Universe (Section 6.3) through the formation of the first nuclei (Section 6.4) and the first neutral atoms (Section 6.5) to the stage at which gravitational clustering gives rise to the large-scale structure that we observe in the present-day Universe (Section 6.6).

At first sight, it might seem impossible to use the models in Chapter 5 to make any predictions about the small-scale behaviour of matter. Cosmological models describe how the scale factor varies with time, but at the present time any change in the scale factor certainly does not have any effect on, for instance, the atoms that make up your body. However one of the major assumptions made in Chapter 5 was that the matter in the Universe is smoothly distributed. This assumption of a uniform distribution of matter is a key to linking the large-scale dynamical behaviour of the Universe to small-scale effects.

To see why this is so, consider a volume of the Universe that, at some particular time, is bounded by an imaginary cube, as shown in Figure 6.1. Let us further suppose that we want to follow the evolution of the matter within this cube at all times using some particular Friedmann–Robertson–Walker model with scale factor $R(t)$. To do this, the edges of the cube must follow the expansion (or contraction) of the model universe.

■ Each edge of the cube has an associated length l. How must the length of each edge change with time if the cube is to follow the expansion or contraction of the model universe?

❏ The length of each edge of the cube must be proportional to the scale factor, i.e. $l \propto R(t)$.

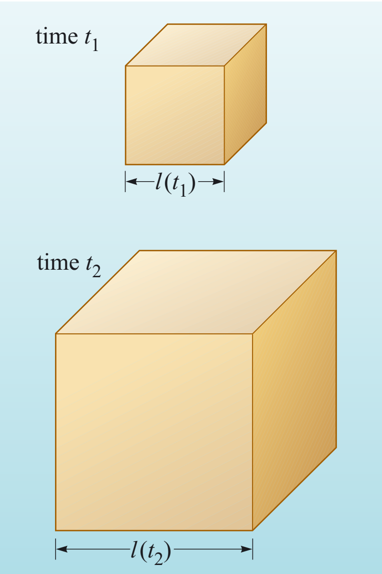


Figure 6.1 An imaginary cubic volume (with sides of length l) that evolves with the expansion of a model universe such that the mass within the volume is constant. Any volume that behaves in this way, whatever its shape, is called a co-moving volume.

Thus the volume ($= l^3$) of the cube at any time t is proportional to $R(t) \times R(t) \times R(t) = (R(t))^3$. This is illustrated by the change in volume shown in Figure 6.1. Although we have chosen to discuss a cubic volume here, a volume of *any* shape that follows the expansion (or contraction) of such a model universe will have a volume $V \propto (R(t))^3$. Such a volume is called a **co-moving volume**.

Because the volume V of a co-moving region, such as the cube, changes as the scale factor changes, but the mass M within it is constant, the density of matter within the co-moving volume, which we denote by ρ_m, also varies with scale factor. In fact,

$$\rho_m = \frac{M}{V} \propto \frac{1}{(R(t))^3} \tag{6.1}$$

Now, the density of matter is an important physical parameter in determining how interactions between particles progress on a microscopic scale. For example, the rate at which molecules in a sample of gas collide with one another increases as the density of the gas is increased. Thus, the large-scale behaviour *is* related to small-scale effects.

This still may not appear to be a great help in understanding the real Universe as opposed to a cosmological model, since we know that the matter in the Universe at the present time does not have a uniform density. Specifically, we saw in Chapter 4 that the distribution of matter is homogeneous only when we consider scales greater than about 200 Mpc: if we look at the Universe on smaller scales, we see large density variations. Thus the average density of the matter in the Universe would seem to be a quantity of limited practical use. However, we shall see later that there is good evidence that at times in the distant past the matter in the Universe was much more smoothly distributed than it is at present – even on relatively small scales. At such times, the average density *does* relate to the small-scale behaviour of matter.

The relationship between density and scale factor that is described by Equation 6.1 holds true for any of the cosmological models described in Chapter 5. The majority of these models are characterized by a scale factor $R = 0$ at time $t = 0$. As was noted in Section 5.3.5 the early expansion phase of any such model is referred to as the *big bang*. Current cosmological evidence strongly favours a model of the Universe that went through a big bang phase – and for the remainder of this chapter we shall only consider big bang models. An immediate consequence that can be noted from Equation 6.1, is that in a big bang model, early stages in the history of the Universe (when $R(t)$ was very small) are characterized by high densities. You may also have noticed that, strictly speaking, the mathematical relationship $\rho_m \propto 1/R^3$ implies an infinite density when $R = 0$. We shall return to consider the significance of such infinite quantities later.

Finally, we should make a brief note about terminology: much of the discussion of this chapter refers to specific times in the history of the Universe. A convention that we adopt throughout the chapter is that the age of the Universe is denoted by t (so, for instance, $t = 1$ s denotes the time at which the Universe was 1 second old). The time now, at which we observe the Universe, is referred to as t_0. So, t_0 represents the current value of the age of the Universe.

6.2 The thermal history of the Universe

A key physical parameter in particle interactions is temperature. It may seem to make no sense to talk of the ‘temperature’ of the Universe. The temperature of matter seems to range from a few degrees above absolute zero within giant molecular clouds, up to temperatures of over 10^7 K that are found in extreme astrophysical environments. However, we will see shortly that there *is* a cosmic temperature; it does not refer to temperature of the *matter* within the Universe, but to the *radiation* that pervades the Universe. This radiation, which is observable today as the cosmic microwave background (CMB), plays an extremely important role in modern cosmology. We shall discuss several aspects of the CMB in this, and in the following chapter, but in this section we shall consider how the existence of this background radiation in an expanding Universe allows us to determine how temperature has changed over cosmic history. The starting point for this discussion is to consider the cosmic microwave background in a little more detail.

6.2.1 The temperature of the background radiation

In Section 5.2.2 you saw that the most significant contribution to the radiation content of the Universe is the cosmic microwave background radiation. The spectral flux density of the CMB peaks at a wavelength of about 1 mm – this is illustrated in the spectrum of the background radiation that is shown in Figure 6.2. The form of this spectrum is highly significant: it is, to a very good approximation, a black-body spectrum – implying it can be associated with a particular temperature.

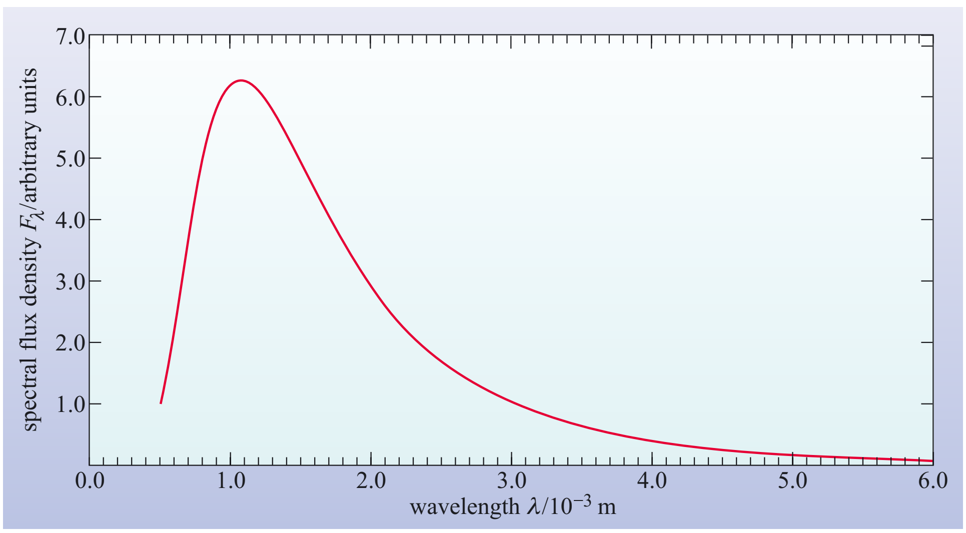


Figure 6.2 The spectrum of the cosmic microwave background. (Note that this spectrum shows the spectral flux density F_λ.)

The characteristic temperature T indicated by any given black-body spectrum is related to the wavelength λ_{peak} at which the spectral flux density (F_λ) is a maximum. According to Wien’s displacement law,

$$(\lambda_{\text{peak}}/\text{m}) = \frac{2.90 \times 10^{-3}}{(T/\text{K})} \tag{6.2}$$

- Calculate the characteristic temperature of the cosmic microwave background radiation.

- Rearranging Equation 6.2

$$(T/\mathrm{K}) = \frac{2.90\times 10^{-3}}{(\lambda_{\mathrm{peak}}/\mathrm{m})}$$

and using $\lambda_{\mathrm{peak}} \approx 1\ \mathrm{mm} = 1 \times 10^{-3}\ \mathrm{m}$

$$(T/\mathrm{K}) = \frac{2.90\times 10^{-3}}{(1\times 10^{-3})} = 2.90$$

So, to one significant figure, the temperature of the cosmic microwave background radiation is 3 K.

Detailed spectral measurements have been used to determine the temperature of the cosmic microwave background to a high degree of accuracy, with a value of $T = 2.725 \pm 0.002$ K being widely accepted.

The fact that the CMB follows a black-body spectrum is, at first sight, puzzling. Black-body spectra are formed when photons are continually absorbed and re-emitted by matter. However, matter in the nearby Universe is transparent to cosmic microwave background photons. Thus there is no interaction between matter and photons, and so nearby matter could not give rise to the observed black-body spectrum. So, if the CMB did form by the interaction of radiation and matter – how could this have occurred? To answer this question we have to consider the effect of the expansion of the Universe on the photons of the cosmic microwave background.

6.2.2 The evolution of the temperature of background radiation

A clue to the origin of the microwave background lies in an effect that was introduced in Chapter 5 – the cosmological red-shift of photons. In Section 5.4.1 you saw that the effect of the expansion of the Universe on a single photon was to increase its wavelength. The relationship between the wavelength λ_0 of a photon that is observed now (when the scale factor has the value $R(t_0)$) and the wavelength λ that the photon had when the scale factor was $R(t)$ is

$$\frac{\lambda}{\lambda_0} = \frac{R(t)}{R(t_0)} \qquad (6.3)$$

Thus, when the scale factor was smaller than it is at present, the wavelengths of photons that are now seen in the cosmic microwave background were all correspondingly smaller. In fact, the background radiation that is now observed as the cosmic *microwave* background, would, when the scale factor was much smaller, have had a peak in another part of the electromagnetic spectrum. For this reason we shall use the term **cosmic background radiation** to denote this radiation at any time in cosmic history. The cosmic microwave background is just the observable form of the cosmic background radiation at the present time.

■ If a microwave background photon currently has $\lambda = 1$ mm, what wavelength would it have had when the scale factor was 1000 times smaller than its present-day value? In which part of the electromagnetic spectrum does this wavelength lie?

❑ Using Equation 6.3, with values of $\lambda_0 = 1$ mm and $R(t)/R(t_0) = 1/1000$ gives

$$\lambda = 10^{-3}\,\text{m}/1000 = 10^{-6}\,\text{m}$$

So when the scale factor was 1000 times smaller than at present, photons that are currently at the peak of the cosmic microwave background had a wavelength of 10^{-6} m, which lies in the infrared part of the spectrum.

Thus, at high redshift, the wavelengths of photons in the cosmic background radiation would have been much shorter than at present, and consequently interactions between photons and matter would have been much more likely. However, before discussing this interaction, we need to consider the form of the red-shifted spectrum in a little more detail.

An important feature of the black-body spectrum is that if the photons that make up such a spectrum are all red-shifted by the same amount, then it will remain a black-body spectrum. Photons that are currently at the wavelength at which the spectrum has a peak, will always be at the peak, but the wavelength of that peak will change. The way in which this wavelength, λ_{peak}, changes with scale factor is given by Equation 6.3. $R(t_0)$ and λ_0 are the current values of $R(t)$ and λ respectively, and so can be considered as constants in Equation 6.3. Thus Equation 6.3 can be written as

$$\lambda_{\text{peak}} \propto R(t) \tag{6.4}$$

However, the temperature of a black-body spectrum is related to the wavelength of the peak of emission by Wien's displacement law (Equation 6.2) which can be rearranged and expressed as

$$T \propto \frac{1}{\lambda_{\text{peak}}} \tag{6.5}$$

Using the relationship between λ_{peak} and the scale factor (Equation 6.4) gives

$$T \propto \frac{1}{R(t)} \tag{6.6}$$

The temperature of the cosmic background radiation at any time is inversely proportional to the scale factor at that time.

This relationship is important because, in principle, it allows us to calculate the temperature of the background radiation at any given epoch for any cosmological model. Remember from Chapter 5 that different cosmological models provide different relationships for the scale factor R as a function of time (see, for example, Figure 5.23).

Even if we do not know the exact way in which the scale factor varies with time, Equation 6.6 shows that if the scale factor was once much smaller than it is at present, then the temperature of the background radiation at that time would have been much higher than it is at present.

- Use Equation 6.6 to express the ratio of the temperature at two times (t_1 and t_2) in terms of the scale factor at those two times.
- Equation 6.6 can be expressed as

$$T(t) = \frac{\text{constant}}{R(t)}$$

So for times t_1 and t_2 we can write

$$T(t_1) = \frac{\text{constant}}{R(t_1)} \quad \text{and} \quad T(t_2) = \frac{\text{constant}}{R(t_2)}$$

respectively. Dividing the first of these equations by the second gives

$$\frac{T(t_1)}{T(t_2)} = \frac{R(t_2)}{R(t_1)}$$

There is, however, a problem in applying Equation 6.6 which is highlighted by the following question.

- What is the predicted temperature of the Universe if the scale factor has a value of zero?
- Since $T \propto 1/R(t)$, if $R = 0$, the predicted temperature would be infinite!

A prediction of an infinite value of any physical quantity is treated with great suspicion by most physicists. Rather than taking this infinite value at face value, it is assumed that our understanding of physical processes is incomplete. The limits at which our knowledge of physical laws break down will be discussed briefly in Section 6.3 and taken up again in Chapter 8. However, for the present discussion, the important point is that at times when the scale factor was very small, the temperature would have been very high.

EXAMPLE 6.1

For the Lemaître cosmological model (in which $\Lambda > \Lambda_E$) and $k = +1$, use Figure 5.23 and Equation 6.6, to sketch a corresponding curve $T(t)$ that shows approximately how the temperature of the cosmic background radiation varies with time.

SOLUTION

The curve that shows how the scale factor R varies with time in the Lemaître cosmological model is shown in Figure 5.23 and is reproduced here as Figure 6.3a.

In order to draw a sketch of how the temperature T varies with time, we need to make use of the relationship between the temperature and the scale factor. This relationship is given by Equation 6.6, $T \propto 1/R(t)$.

The question asks for a *sketch* of how T varies with time. The implication of this is that the curve that shows $T(t)$ does not have to be exact, but that it should show the most important features of how the temperature varies with time. A way of doing this is to consider a few times (labelled A, B, C and D) on the corresponding curve of $R(t)$ as shown in Figure 6.3a. At each time, we shall use Equation 6.6 to deduce how T is behaving and use this information to help us draw a sketch of $T(t)$. The deductions that can be made about T at these times are shown in Table 6.1.

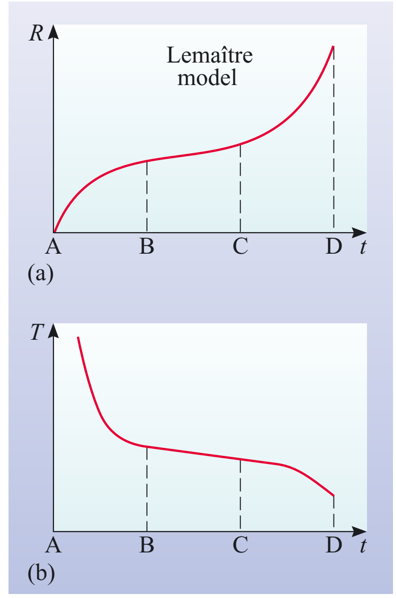


Figure 6.3 (a) Scale factor, and (b) temperature as functions of time for the Lemaître cosmological model. A, B, C and D are times that are referred to in Table 6.1.

Table 6.1 The behaviour of the scale factor R at various times indicated on Figure 6.3a and the inferred behaviour of the temperature T at those times.

Time	Behaviour of R at this time	Behaviour of T at this time
A	$R = 0$	$T = 1/R = \infty$
B	R has increased to some value and now does not vary much with time	T must decrease to some value and also only change slowly with time
C	R has a value that is slightly higher than that at B	T must have a value that is slightly lower than that at B
D	R is increasing to very high values	T must decrease to very low values

We can now use the deductions about the way in which T varies at these times to draw a sketch. Starting with time A, we clearly cannot plot an infinite temperature at A, so we simply show T as having a very high value as we approach $t = 0$. At time B, we simply choose a finite value of temperature, and note that the temperature at time C is slightly lower than at B. Finally, at time D, the temperature decreases to very low values. These points are shown on Figure 6.3b, and the final stage is to draw a smooth curve through these points to complete the sketch.

QUESTION 6.1

For all the Friedmann–Robertson–Walker models with $k = 0$, shown in Figure 5.23, use Equation 6.6 to draw a corresponding curve $T(t)$ that shows approximately how the temperature of the cosmic background radiation would vary with time.

The answer to Question 6.1 illustrates the point that in any big bang model (i.e. one that has $R = 0$ at $t = 0$) the temperature of the cosmic background radiation would have been very high in its early stages. Such a scenario is often referred to as the **hot big bang**.

The change in wavelength of the cosmic microwave background has a profound effect on the way in which photons interact with matter. At times when the temperature of the cosmic background radiation was very high, the typical photon energy would have been greater than the ionization energy of the hydrogen atom. Under these conditions, the baryonic matter in the Universe would have been in the form of a plasma.

The *opacity* of a medium is a measure of the extent to which the medium is opaque to radiation.

- What is the qualitative difference between the opacity of a plasma and that of an un-ionized gas?

- A plasma tends to have a much higher opacity than an un-ionized gas, i.e. un-ionized gases tend to be much more transparent than plasmas.

The reason for the dramatic difference in opacity between a plasma and an un-ionized gas is the presence of free electrons in the plasma. Photons interact with a plasma primarily by scattering from the free electrons in the plasma (Figure 6.4 – this process is called *Thomson scattering* after the discoverer of the electron J. J. Thomson). The degree of interaction between photons and electrons in a plasma can be very high, and this offers a clue as to the origin of the near perfect black-body spectrum of the background radiation. The conditions for forming a black-body spectrum are that there must be many collisions between the material that makes up a thermal source and the photons that are radiated by it. So an interpretation of the black-body spectrum of the cosmic microwave background is that it was formed at a time when the Universe consisted of a hot plasma, and so there were many collisions between the photons and the free electrons. As the Universe has expanded, the wavelengths of the photons have increased, and the black-body spectrum has shifted to longer wavelengths. Consequently, the temperature associated with this black-body spectrum has dropped with the expansion of the Universe.

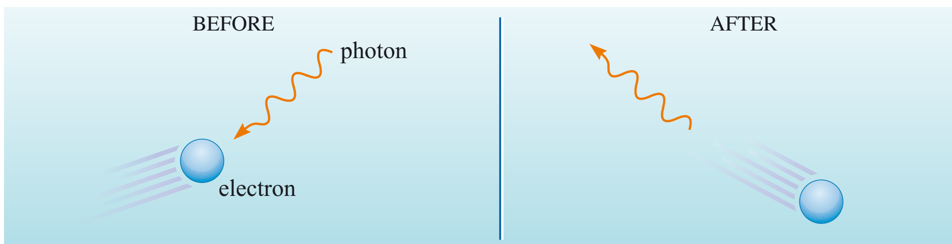


Figure 6.4 The interaction between photons and free electrons in a plasma.

6.2.3 The evolution of energy densities in the Universe

So far, we have only considered one physical property of the cosmic background radiation – its temperature. However, to establish whether this background radiation plays an important role in the evolution of the Universe, it is necessary to consider its *energy density* and how this quantity varies with scale factor.

Recall that in Chapter 5, the behaviour of cosmological models was shown to depend on the density of matter ρ_m, and on the cosmological constant Λ. In the models that we considered there, we simply assumed that electromagnetic radiation was a minor constituent of the Universe. However, we now want to question this assumption – so we must compare the importance of these three components: matter, cosmological constant, and electromagnetic radiation. The physical parameter that determines the importance of any one of these components within a cosmological model is the energy density, i.e. the energy per unit volume due to that component. If the energy density of any one component far exceeds that of the other two, then it is this component that will have the dominant effect on the dynamical behaviour of the Universe.

So, let us now consider the current energy densities due to radiation u_r, matter u_m and the cosmological constant u_Λ.

The energy density of the cosmic microwave background is the total energy of all the microwave background photons per cubic metre of space. Note that the energy density (like the mass density) is defined per cubic metre, i.e. for a physical volume of space that is *not* co-moving. In an expanding universe, the energy density can therefore be expected to decrease as the universe expands.

■ What are the SI units in which energy density should be expressed?

❑ Since the SI unit of energy is the joule, and the unit of volume is m^3, the SI unit of energy density is $J\,m^{-3}$.

The current energy density of the cosmic background radiation can be found from measurements of the CMB and has a value of $u_{r,0} \approx 5 \times 10^{-14}\,J\,m^{-3}$. ($u_{r,0}$ is a shorthand way of writing $u_r(t_0)$ – the value of u_r at the present time.)

The energy density of matter $u_{m,0}$ can be found from density of matter in a straightforward way as the following question illustrates.

QUESTION 6.2

The current average *mass density* of all matter, both luminous and dark, is estimated to be about $\rho_{m,0} \approx 3 \times 10^{-27}\,kg\,m^{-3}$. By using the equivalence between energy E and mass m given by $E = mc^2$, calculate the current average *energy density* due to matter.

The answer to Question 6.2 shows that at the present time, the energy density of matter is $u_{m,0} \approx 3 \times 10^{-10}\,J\,m^{-3}$. Thus at the present time, the energy density due to matter exceeds the energy density in the cosmic microwave background by a factor of several thousand.

Finally, we consider the energy density due to the cosmological constant. In Chapter 5, it was noted that the cosmological constant Λ has an associated density

$$\rho_\Lambda = \Lambda c^2/8\pi G \qquad (6.7)$$

■ Give an expression for the *energy* density (u_Λ) of the vacuum in terms of Λ.

❑ The energy density of the vacuum is obtained by multiplying Equation 6.7 by c^2,

$$u_\Lambda = \rho_\Lambda c^2$$

and so

$$u_\Lambda = \Lambda c^4/8\pi G \qquad (6.8)$$

As was noted in Chapter 5, the energy that may be associated with the cosmological constant is often referred to as *dark energy*. Consequently the quantity u_Λ can be interpreted as the energy density of dark energy. The nature of this dark energy is a mystery, but recent observations imply that u_Λ has a value of about $9 \times 10^{-10}\,J\,m^{-3}$. So, rather surprisingly, dark energy makes the dominant contribution to the total energy density of the Universe at the present time.

It might seem then that the cosmic background radiation is an insignificant component of the total energy density of the Universe. However, this was not always the case. To see why, it is necessary to compare the way in which the three energy densities u_r, u_m and u_Λ change with scale factor.

We start with the simplest case of the three, which is the energy density of the dark energy. By inspecting the terms on the right-hand side of Equation 6.8 we can see that this energy density depends only on values of physical constants (c, G and Λ). Thus, this energy density does not change with scale factor. As noted above, u_Λ currently has a value of about $9 \times 10^{-10}\,\mathrm{J\,m^{-3}}$, and this value has been constant throughout the history of the Universe (with perhaps one brief, but important, exception that we shall discuss later). However, the fact that u_Λ is constant and relatively large, does not mean that it has always been the most important factor in determining how the Universe evolves.

Next, let's consider the how the energy density due to matter changes with scale factor. This is found from the (normal) density of matter.

■ For a cosmological model in which matter is uniformly distributed, write down an equation that describes how the density of matter changes with scale factor.

❑ We have already seen that the density of matter $\rho_m(t)$ varies according to Equation 6.1

$$\rho_m(t) \propto \frac{1}{R(t)^3} \tag{6.1}$$

The energy density of matter is related to the density of matter by $u_m = \rho_m c^2$, but c is a constant, so we can write

$$u_m(t) \propto \frac{1}{R(t)^3} \tag{6.9}$$

The behaviour of the energy density of *radiation* can be analysed by taking a similar approach to that taken when we examined the way in which the density of matter changes with scale factor. The first step is to consider the number of photons per cubic metre, i.e. the *number density* of photons $n(t)$. Assuming that cosmic background photons are neither created nor destroyed during the relevant part of the expansion we can expect that

$$n(t) \propto \frac{1}{R(t)^3} \tag{6.10}$$

So the number density of cosmic background photons behaves in a similar way to the density of matter. But what about the energy density? Here there is a difference. The energy of a photon of frequency f is given by

$$\varepsilon_{ph} = hf \tag{6.11}$$

where h is the Planck constant. The frequency f and wavelength λ of electromagnetic radiation are always related by $f = c/\lambda$, so we can also say

$$\varepsilon_{\mathrm{ph}} = \frac{hc}{\lambda} \tag{6.12}$$

But as the Universe expands (i.e. as $R(t)$ increases) the wavelength of a photon will also increase (Figure 6.5). The photon wavelength is proportional to the scale factor

$$\lambda \propto R(t) \tag{6.13}$$

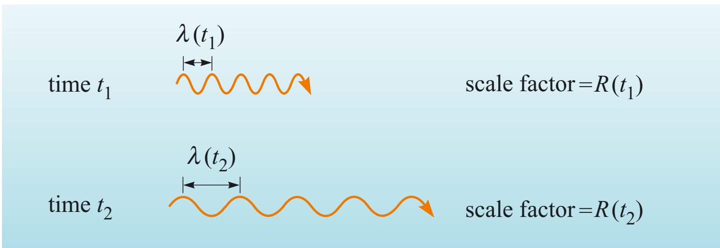


Figure 6.5 As the Universe expands (from scale factor $R(t_1)$ to $R(t_2)$), the wavelength of any photon will increase in proportion to the scale factor.

Hence for each photon in the cosmic background radiation

$$\varepsilon_{\mathrm{ph}} \propto \frac{1}{R(t)} \tag{6.14}$$

Now, the energy density of radiation $u_{\mathrm{r}}(t)$ is given at any time t by

$$u_{\mathrm{r}}(t) = n(t) \times \varepsilon_{\mathrm{ph}}(t) \tag{6.15}$$

It follows that

$$u_{\mathrm{r}}(t) \propto \frac{1}{R(t)^4} \tag{6.16}$$

Comparing this with Equation 6.9 shows that the energy density of radiation behaves differently from the energy density of matter. Specifically, the energy density of radiation is inversely proportional to the *fourth* power of the scale factor, whereas the energy density of matter is inversely proportional to the *third* power of the scale factor.

Figure 6.6 shows how all three energy densities (u_Λ, u_{m}, and u_{r}) vary as a function of scale factor. At the present, the energy densities of the dark energy ($9 \times 10^{-10}\,\mathrm{J\,m^{-3}}$) and of matter ($3 \times 10^{-10}\,\mathrm{J\,m^{-3}}$) are far in excess of the energy density of radiation ($5 \times 10^{-14}\,\mathrm{J\,m^{-3}}$). As we look back to earlier times however, when the value of the scale factor was smaller, we can see that both the energy densities of radiation and of matter were greater in the past than at present. When the value of $R(t)/R(t_0)$ was less than about 0.1, both the energy density of matter and of radiation exceeded the energy density of the dark energy.

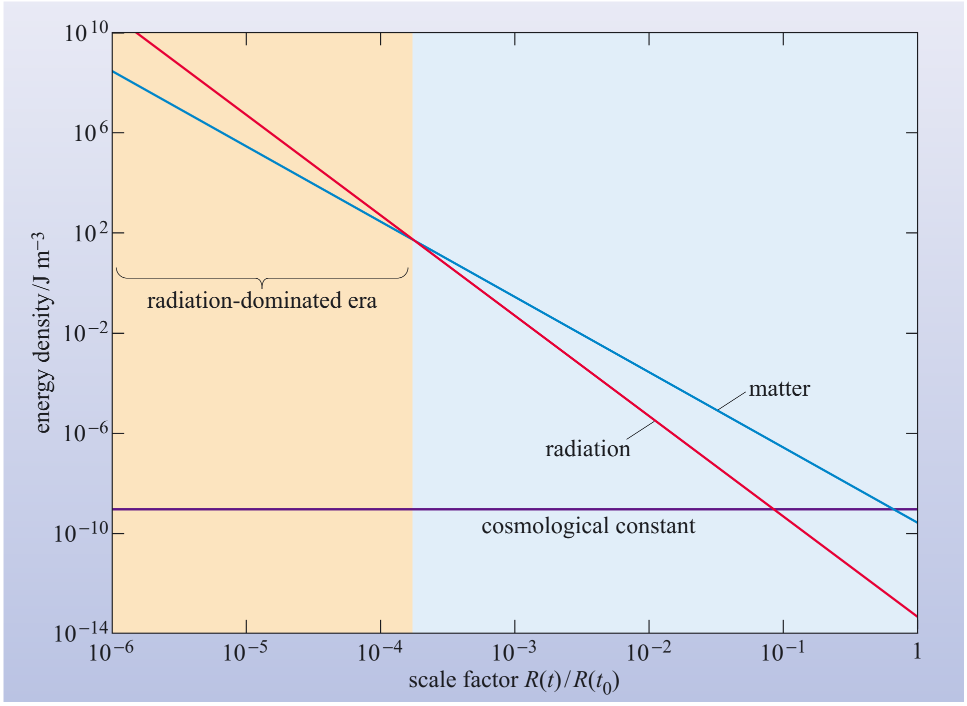


Figure 6.6 The energy densities of matter (blue line) and radiation (red line) as a function of scale factor. At a time when $R(t)/R(t_0) \approx 10^{-4}$ the energy densities of matter and radiation were equal. Prior to this time, the energy density of radiation exceeded that of matter – during this era the dynamical evolution of the Universe was determined by its radiation content. After this time, the energy density of matter was greater, so it was the matter in the Universe that controlled its dynamical evolution. The behaviour of the energy density due to the cosmological constant is also shown (purple line) – this does not vary with redshift and is exceeded by the energy densities in matter and radiation at early times.

Furthermore, as shown in Figure 6.6, the energy density of radiation has declined more rapidly than the energy density of matter. Indeed there was a time when the value of $R(t)/R(t_0)$ was such that these two energy densities were equal. This appears to have occurred when $R(t)/R(t_0) \approx 10^{-4}$. For most plausible cosmological models this corresponds to a time when the age of the Universe was a few times 10^4 years. As we look back to even earlier times, when $R(t)/R(t_0)$ was even smaller, we see that the energy density of radiation exceeded that of matter – this period of the history of the Universe is called the **radiation-dominated era**.

The key points of this discussion so far can be summarized as follows:

1 At the present time, the energy density due to radiation is much lower than the energy density of matter or the energy density of dark energy.

2 As the Universe expanded, the energy density of radiation decreased more rapidly than the energy density of matter. The energy density of dark energy has remained constant with time.

3 At times when the scale factor was less than about 10^{-4} of its current value, the energy density of radiation would have exceeded the energy density due to matter. At this time, the energy density of dark energy would have been negligible in comparison to the energy densities of radiation or matter.

As you will shortly see, the existence of an early radiation-dominated era has a profound effect on the dynamical evolution of the Universe.

At this point you may be wondering how is it that the energy density of matter and radiation change in different ways? The numbers of photons and particles within a co-moving volume remain constant, so this cannot be the origin of the difference. The answer lies in the fact that the energies of photons change with the expansion of the Universe, whereas the masses (and hence energies) of particles such as protons and electrons and of any cold-dark matter particles remain constant.

6.2.4 A radiation-dominated model of the Universe

We have just seen that in the early Universe, the dominant energy density is that due to the radiation within the Universe. The Friedmann equation that was described in Chapter 5 (Box 5.4) can be solved for such conditions and the way in which the scale factor varies with time for such a model is shown in Figure 6.7. One important feature of such a model is that the scale factor varies in the following way:

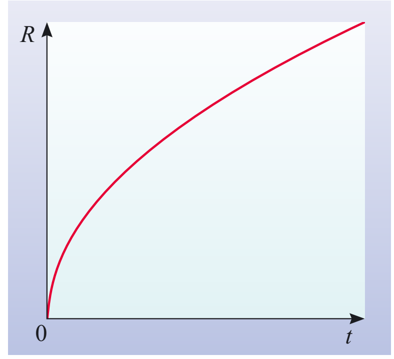


Figure 6.7 The evolution of the scale factor with time in a cosmological model in which the dominant contribution to the energy density arises from the radiation within the Universe (i.e. during the radiation-dominated era).

$$R(t) \propto t^{1/2} \tag{6.17}$$

Because the energy density of radiation is dominant for times when $R(t)/R(t_0) < 10^{-4}$, all cosmological models which start at $t = 0$ with $R(0) = 0$, will go through a phase that is well described by this radiation-dominated model. Thus we are in the rather remarkable position that regardless of which type of cosmological model best describes the Universe at the present, we can be reasonably confident that we know how the scale factor varied with time in the first few tens of thousands of years of the big bang.

However, the temperature of the background radiation varies with scale factor according to $T(t) \propto 1/R(t)$ (Equation 6.6). It follows that during the radiation-dominated era the temperature of the background radiation varies with time according to

$$T(t) \propto t^{-1/2} \tag{6.18}$$

This describes how temperature changes with time in an expanding universe where the energy density of radiation is the dominant component. Of course, to use Equation 6.18 to predict the temperature, it is necessary to know the constant of proportionality between T and $t^{-1/2}$. In fact, this can be derived from the Friedmann equation and Equation 6.18 becomes

$$(T/\mathrm{K}) \approx 1.5 \times 10^{10} \times (t/\mathrm{s})^{-1/2} \tag{6.19}$$

(Note that T is measured in kelvin and t in seconds.) Equation 6.19 is an approximate relationship. As you will see later, other physical processes can change the temperature of the radiation in the real Universe during the radiation-dominated phase of its expansion.

- ■ What is the temperature when the age of the Universe is one second?
- ❑ By substituting a value of $t = 1$ s in Equation 6.19 the temperature is 1.5×10^{10} K.

Thus the temperature of the Universe in the first few seconds of the big bang was higher than the highest temperatures that are found in the cores of the most massive stars (where temperatures may reach about 10^9 K). This immediately suggests that nuclear reactions may have occurred in any matter that was present at this time. We will look into such processes in more detail in Section 6.4, but in the next section we will discuss even more extreme conditions: we will consider the processes that occurred when the Universe was less than 1 second old.

QUESTION 6.3

Rearrange Equation 6.19 to find an expression for the age of the Universe for a given temperature. What was the age of the Universe when the temperature was 10^6 K? Express your answer in terms of years.

6.3 The early Universe

We have seen that when considered together, the cosmic microwave background and the expansion of the Universe imply that there was an early phase of the history of the Universe which was characterized by high temperatures and high densities. In particular, the temperature at any time t can be estimated using Equation 6.19. A natural question to ask then is how far back towards $t = 0$ can we go in understanding processes in the Universe? This will be the first major question that we shall consider in this section, and we will find that there is a limit to our knowledge of the evolution of the Universe. The remainder of the section will be concerned with understanding the processes that occurred before the Universe was about 1 second old.

6.3.1 Cosmology and the limits of physical theory

We have seen that the Friedmann equation gives a model for a radiation-dominated Universe that is characterized by the scale factor having a value of zero at the instant of $t = 0$. As we said in Section 6.2.2, the naive interpretation of this is that the Universe came into existence with an infinitely high temperature; the truth of the matter is that we don't really understand the physical processes in the very early Universe. So, how early in the history of the Universe can we be confident that our physical theories really do apply? There are essentially two answers to this question, which reflect two levels of certainty in physical theory. The first approach is to say that theories are only well-tested for the ranges of physical conditions that can be explored by experiments. Thus, we may have a good deal of confidence in describing the Universe at times when the particle energies were similar to the highest values that can be imparted in large accelerator experiments. At present, this limit corresponds to being able to describe physical processes in the Universe that occurred after the temperature fell to below 10^{15} K, which corresponds to a time of $t \sim 10^{-9}$ s.

An alternative approach is to apply physical theories to conditions that never have been, and probably never will be, tested in the Earth-bound laboratory and to look for observable consequences in nature. Clearly, this is a somewhat more speculative approach than having to rely on 'tried-and-tested' physical theory. However, it is one way in which physical theories can be explored and developed, and is a very exciting field for cosmologists.

While it might be expected that physical theories could be extrapolated to describe processes at ever increasing temperatures, it turns out that there is a well recognized limit to our theoretical understanding of the processes of nature. This limit arises because of a surprising incompatibility between the physical theory that is used to describe the interactions of subatomic particles and the theory that describes gravity. The interactions of subatomic particles are described by a branch of quantum physics called the **standard model** of elementary particles. The gravitational interaction is described, as you saw in Chapter 5, by Einstein's general theory of relativity.

The general theory of relativity describes effects that were not explained by Newton's theory of gravity, and as far as the theory can be tested, there have been no observations or measurements to suggest that the theory is incorrect. The standard model of elementary particles is much more amenable to being tested by experiment than is general relativity, and its predictions have been well tested by laboratory measurements. Despite the fact that both theories appear to be sound, it has proven impossible to join them together to form a single consistent theory.

Thus physicists expect that neither general relativity nor the standard model offers a full description of the fundamental interactions of nature, and propose that there must be a unifying 'theory of everything' that is yet to be discovered. In particular, such a theory is needed to describe processes in the very extreme conditions that occurred when the Universe was less than about 10^{-43} s old. This limiting time is called the **Planck time** and represents the limit of how far back in time towards $t = 0$ can be investigated using current physical theory. (The Planck time is $t_{\text{Planck}} = (Gh/2\pi c^5)^{1/2} = 5.38 \times 10^{-44}$ s.)

6.3.2 Conditions and processes in the early Universe

To set the scene for our account of the evolution of the Universe from a time of about 10^{-43} s, it is necessary to review some important physical concepts and processes. A key feature of the early Universe is that the radiation and matter were interacting so much that they were in a state of **thermal equilibrium**. This means the temperature of matter (as defined by the distribution of particle energies) and the temperature of radiation (as defined by the black-body spectrum) were equal.

At the high temperatures that existed in the early Universe, the composition of the Universe, in terms of the particles that were present, was determined by the typical energy that was available in particle interactions. This energy is termed the **interaction energy**, and is related to the temperature by

$$E \sim kT \tag{6.20}$$

where k is the Boltzmann constant ($k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$).

(The '~' sign is used to indicate a very approximate relationship. Note also that some books use $E \sim 3kT$, rather than Equation 6.20 given here.)

Note that it is common practice to express the interaction energy in terms of electronvolts (eV) where $1\,\mathrm{eV} = 1.60 \times 10^{-19}\,\mathrm{J}$. The energies involved are usually expressed in MeV ($1\,\mathrm{MeV} = 10^6\,\mathrm{eV}$) or GeV ($1\,\mathrm{GeV} = 10^9\,\mathrm{eV}$).

■ Calculate the interaction energy when the temperature is 10^{14} K. Express your answer in joules and GeV.

❑ The interaction energy is found using Equation 6.20

$$E \sim kT = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}} \times 10^{14}\,\mathrm{K} = 1.38 \times 10^{-9}\,\mathrm{J}$$

In terms of GeV

$$E = 1.38 \times 10^{-9}\,\mathrm{J}/(1.60 \times 10^{-19}\,\mathrm{J\,eV^{-1}}) = 8.63 \times 10^{9}\,\mathrm{eV} = 8.63\,\mathrm{GeV}$$

So the interaction energy is 9 GeV (to one significant figure).

The fundamental interactions and their evolution

A key idea in the cosmology of the very early Universe relates to the four *fundamental interactions* of nature: these are gravitational, electromagnetic, weak and strong interactions (see Box 6.1). In the context of the present-day Universe, these interactions seem quite distinct. They operate over quite different ranges – the weak and strong interactions act only over distances that are comparable to the size of an atomic nucleus. Furthermore, the 'strength' of these interactions can be defined in a way that allows sensible comparison. The weakest interaction is gravity, and then in ascending order of 'strength', there follows the weak interaction, the electromagnetic interaction, and appropriately enough, the strong interaction.

BOX 6.1 FOUR FUNDAMENTAL INTERACTIONS

At a fundamental level nature has just four types of interaction. This means that any physical process, for example, the scattering of a photon off an electron, or the generation of electricity in a nuclear power plant, can be analysed in terms of one or more of these four interactions.

The fundamental interactions are:

1. The *gravitational interaction*, which for instance, keeps the Earth in orbit around the Sun. This acts over large distances, so it is part of our everyday experience.
2. The *electromagnetic interaction*, which for instance, keeps electrons bound to atoms. Like the gravitational interaction, this interaction also acts over long distances.
3. The *strong interaction*. This interaction only acts over distances comparable to the diameter of a nucleus. An example of the effect of the strong interaction is the binding of the protons and neutrons together in the nucleus of an atom. The strong interaction overcomes the mutual repulsion that acts between the positively charged protons in a nucleus.
4. The *weak interaction*. This is also a short-range interaction, that acts only on scales comparable to that of the nucleus. An example of the effect of the weak interaction occurs in the transformation of a neutron to a proton in β^--decay.

The standard model of elementary particles explains the operation of the strong, weak and electromagnetic interactions in terms of so-called 'exchange particles' that carry energy and momentum between interacting particles and thereby account for the action of a 'force' in a fundamental way. The exchange particles for the various interactions are as follows:

- The *photon*: the exchange particle of the electromagnetic interaction.
- The *W^+, W^- and Z^0 bosons*: the exchange particles of the weak interaction.
  The masses of these particles are responsible for the short range of the weak interaction.
- The *gluons*: a family of eight similar particles that are responsible for the strong interaction. These particles are confined within the protons or neutrons that comprise a nucleus.

The gravitational interaction is described by the general theory of relativity. Within this theory gravity arises from the curvature of space–time rather than from a particle interaction.

It is suspected that all four interactions may be different manifestations of a single fundamental type of interaction. The reason for such a belief is partly philosophical and partly experimental. The 'philosophical' justification for the unification of interactions is that this type of approach – reducing the physical world to what appears to be the minimum number of particles and processes – has been outstandingly successful, and physicists see this as the next logical advance. If this sounds wildly idealistic, then the 'experimental' justification should offer some

reassurance. A key idea in demonstrating that two interactions are linked is that under certain physical conditions they should behave in the same way. So, for instance, the strength of two interactions may become the same.

Experiments using particle accelerators have revealed that the strength of interactions depends on the interaction energy. In particular, the strengths of the electromagnetic and weak interactions are observed to become closer to one another at high interaction energies. At interaction energies of about 1000 GeV, the strengths of these two interactions are predicted to be the same, and the electromagnetic and weak interactions should appear as different manifestations of a single underlying *electroweak* interaction.

It is believed that the unification of the other interactions occurs at very much higher interaction energies. The unification of the strong interaction with the electroweak interaction – which is termed 'grand unification' is predicted to occur at an interaction energy of about 10^{15} GeV. The theoretical framework that is used to describe this unified interaction is called a **grand unified theory** or **GUT**.

QUESTION 6.4

Calculate the temperature corresponding to the minimum interaction energy required for grand unification. Hence calculate the age of the Universe when the strong and electroweak interactions became distinct.

(It is appropriate to quote the results as order-of-magnitude estimates, i.e. to the nearest whole number power of ten.)

The energy at which the gravitational interaction might become unified with the other interactions, if such a thing happens at all, is expected to be higher still – about 10^{19} GeV. At such extreme interaction energies the gravitational interaction might become important for interactions between particles (at lower energies, the gravitational interaction has a negligible effect on particle interactions). In terms of the evolution of the Universe, an interaction energy of 10^{19} GeV corresponds to the Planck time ($\sim 10^{-43}$ s). As has already been mentioned, there is no accepted 'theory of everything' which allows the processes that occurred in this **Planck era** to be understood.

The interaction energies associated with GUT interactions and the Planck era are extreme – there is probably no environment in the present-day Universe in which particles interact with such energy. Thus, it is unlikely that direct experimental verification will ever be made of theories that describe interactions at such high energies.

The expected behaviour of the fundamental interactions over the first few moments of the history of the Universe can be summarized as follows. Prior to $t \sim 10^{-43}$ s all four fundamental interactions may have been unified. After this time, the gravitational interaction became distinct from the GUT interaction. Some time later, at $t \sim 10^{-36}$ s (see the answer to Question 6.4) the strong interaction and the electroweak interaction became distinct from one another. Finally, at $t \sim 10^{-12}$ s, when the typical interaction was about 1000 GeV, the weak and electromagnetic interactions took on the form in which they act in the present-day Universe. This evolution of the fundamental interactions is illustrated schematically in Figure 6.8.

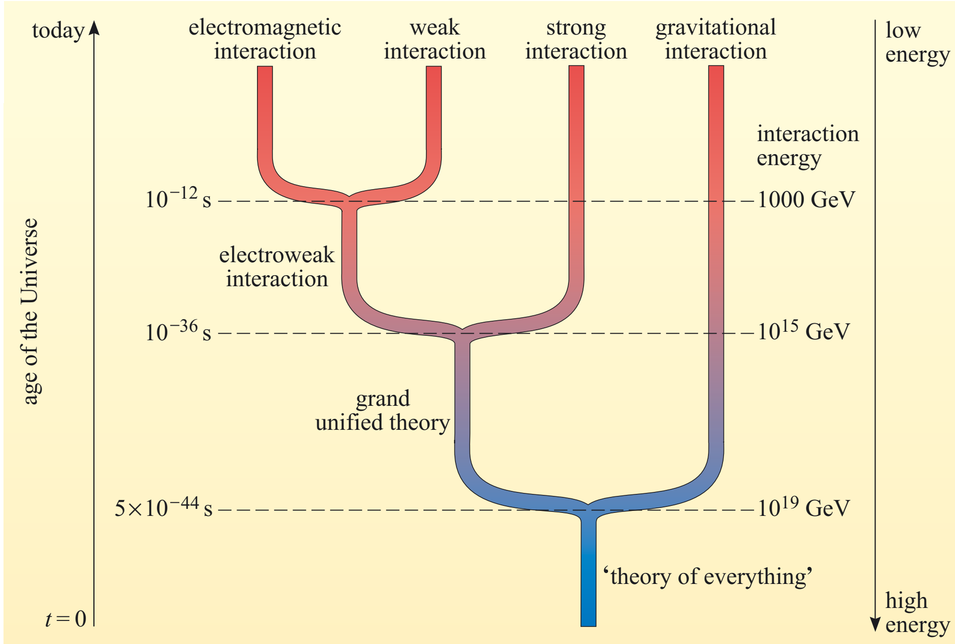


Figure 6.8 The evolution of the fundamental interactions with cosmic time.

Note that the earliest 'branch' shown in this diagram represents the end of the Planck era. The second branch corresponds to the time that grand unification ends – as we shall see later, it is speculated that this change is associated with a dramatic cosmological event, known as *inflation*. The third branch is associated with the separation of the electroweak interaction into the weak and electromagnetic interactions. This is the only one of the three branches that can currently be regarded as experimentally supported; the others are still very speculative.

Particle–antiparticle pair creation

An important process that occurs when interaction energies become very high is that particles and antiparticles can spontaneously form in pairs. This process is known as **pair-creation**, and it has an important effect on the composition of the Universe at early times. Interactions obey a set of conservation rules: the conserved quantities include energy, electric charge, and baryon and lepton numbers (See Box 6.2). The energy available in an interaction plays a vital role in determining which particles may form. Provided that all other conservation rules are obeyed, a particle that has a mass m can be formed if the available energy is equal to or exceeds its *mass energy*, which is given by $E = mc^2$.

For example, when the age of the Universe was 10^{-12} s (i.e. about the time at which electroweak unification ended) the temperature was 10^{16} K and the typical energy of a photon or particle was about 10^3 GeV. Thus any interactions that occurred could easily supply 10^3 GeV to create a new particle. This energy exceeds the mass energy of all of the quarks and the leptons (see Box 6.2). As a result, the material content of the Universe at this time includes all types of lepton and quark and their respective antiparticles.

BOX 6.2 QUARKS AND LEPTONS

The ultimate building blocks of matter are two families of fundamental particles called quarks and leptons. There appear to be six members of each family, as shown in Tables 6.2 and 6.3. Note that for each fundamental particle shown here there exists a corresponding antiparticle with the opposite charge, but the same mass.

Table 6.2 The six quarks.

	Name	Symbol	mass $\times c^2$/GeV
quarks with electric charge of $+2e/3$	up	u	5×10^{-3}
	charm	c	1.5
	top	t	1.8×10^{2}
quarks with electric charge of $-1e/3$	down	d	8×10^{-3}
	strange	s	0.16
	bottom	b	4.25

Table 6.3 The six leptons.

	Name	Symbol	mass $\times c^2$/GeV
leptons with electric charge $-e$	electron	e^-	5.11×10^{-4}
	muon	μ^-	0.106
	tauon	τ^-	1.78
leptons with zero electric charge	electron neutrino	ν_e	$< 1.5 \times 10^{-9}$
	muon neutrino	ν_μ	$< 1.7 \times 10^{-4}$
	tauon neutrino	ν_τ	$< 2.4 \times 10^{-2}$

Quarks are the fundamental particles that make up baryons, such as the proton and the neutron. A proton comprises two up quarks and a down quark, while a neutron comprises two down quarks and an up quark. Quarks are never found in isolation in laboratory experiments – they are always confined in clusters consisting of three quarks or three antiquarks, or in a pair comprising a quark and antiquark. The particles formed by such combinations of quarks are generally termed **hadrons**. A hadron that consists of three quarks is a **baryon**, and a hadron that comprises three antiquarks is an **antibaryon**. A quantity that is conserved in all known particle interactions is the **baryon number** – the baryon number of each quark is 1/3 while that of each antiquark is –1/3. So the baryon number of a baryon is +1, and that of an antibaryon is –1. The baryon number of all other particles is 0.

The family of **leptons** includes the electron (e^-) and two other charged particles: the muon (μ^-) and the tauon (τ^-). The other members of the lepton family are the three types of neutrino – there is one type of neutrino for each of the charged leptons (ν_e, ν_μ, ν_τ). As in the case of quarks, for each type of lepton there exists a corresponding antilepton. Unlike quarks, leptons are not confined and *can* be found in isolation in laboratory experiments.

In interactions that involve leptons, there is a conserved quantity called the **lepton number**. The lepton number of each of the leptons shown in Table 6.3 is +1 and that of each of the antileptons is –1. The lepton number of all other particles is 0.

QUESTION 6.5

Consider the following reaction in which a free neutron decays into a proton, an electron and an electron antineutrino ($\overline{\nu}_e$) (such a reaction is an example of β^--decay).

$$n \rightarrow p + e^- + \overline{\nu}_e$$

(a) What is the baryon number (i) before, and (ii) after this process? Hence show that baryon number is conserved.

(b) What is the lepton number (i) before, and (ii) after this process? Hence show that lepton number is conserved.

(c) What combination of quarks constitutes (i) a neutron, and (ii) a proton? Hence express β^--decay as a reaction involving quarks and leptons only.

The effect of the other conservation rules in determining which particles may be formed in an interaction is profound. Of particular importance are the conserved quantities known as total baryon number and total lepton number (described in Box 6.2). Consider a simple interaction in which two energetic photons interact to form particles

$$\gamma + \gamma \rightarrow \text{‘particles’}$$

- What is (a) the total lepton number of the two photons; (b) the total baryon number of the two photons?

- Photons have a lepton number of zero and a baryon number of zero. Thus (a) the total lepton number of the two photons is zero, and (b) the total baryon number of the two photons is zero.

Since lepton and baryon number are conserved, the two-photon reaction can only form products whose total lepton and baryon number is zero. This does not mean that the lepton and baryon number of each particle that is formed must be zero, but that the *sum* of the lepton and baryon numbers for all the particles that are formed must be zero. For instance, the reaction may result in the production of an electron (lepton number +1) and a positron (an antielectron, lepton number −1)

$$\gamma + \gamma \rightarrow e^+ + e^- \tag{6.21a}$$

Since electrons and positrons have a baryon number of zero, this reaction clearly conserves baryon number.

The pair-creation reaction described by Equation 6.21a is reversible – a positron and an electron can combine to produce two photons according to

$$e^+ + e^- \rightarrow \gamma + \gamma \tag{6.21b}$$

This process, in which a particle and its corresponding antiparticle interact and disappear is called **annihilation**. This process can occur for any particle–antiparticle pair, and the total energy of the photons can be found using the mass energy equivalence relation ($E = mc^2$).

QUESTION 6.6

Calculate the minimum interaction energy required for electron–positron pair production (Equation 6.21a). Express your answer in electronvolts. At what temperature is electron–positron pair production likely to occur?

Note that given sufficiently energetic photons, a two-photon reaction could generate any lepton–antilepton pair. Similarly, a two-photon reaction could generate a quark–antiquark pair

$$\gamma + \gamma \rightarrow \text{q} + \overline{\text{q}} \tag{6.22}$$

(Where the symbol $\overline{\text{q}}$ represents an antiquark.) The photon–photon interaction is just one of many types of interaction that could occur, but without going into detail about these, we can see that the conservation rules will dictate that, provided a sufficiently high interaction energy is available, the Universe will be populated by a mixture of quarks and antiquarks and of leptons and antileptons.

We have concentrated here on quarks and leptons, but there are other particles too that were present in the early Universe. There are two categories of particles that deserve mention. The first are the exchange particles that act to transmit the fundamental interactions of nature (in the parlance of particle physics, these particles 'mediate' the interactions). The most familiar of these is the photon – a massless particle that mediates the electromagnetic interaction. In addition, there are other particles, as described in Box 6.1, that mediate the strong and weak interactions.

From an astronomical point of view, the other important category of particle comprises the massive, stable particle (or particles) that make up dark matter. As has been mentioned, the nature of dark matter is not known, but it is believed to be in the form of particles that are neither baryons or leptons. Presumably, such particles must have been present in the early Universe, but until we have a better idea of what they are, their origin remains a mystery. As far as this chapter is concerned, we shall assume that there are dark matter particles present in the early Universe, but we shall also assume they are essentially non-interacting, so we shall not need to mention them. We will however, consider the role of dark matter at later times – when structure begins to form as a result of gravitational collapse.

Having now reviewed some of the important processes in the early Universe, we can begin a chronological account of the evolution of the Universe. Throughout the following discussion we shall indicate the time t, and the temperature T and interaction energy E at these times (note that in most cases, these are order of magnitude estimates only).

So, let's start at (almost) the very beginning.

6.3.3 The Planck era

$t < 5 \times 10^{-44}$ s, $T > 10^{32}$ K, $E > 10^{19}$ GeV

We have already noted that there is no physical theory to describe processes of the Planck era ($t < 5 \times 10^{-44}$ s). When the age of the Universe was less than the Planck time, it is believed that the fundamental interactions would have had similar strengths, but without a consistent physical theory very little can be predicted about

what would happen at this time. We shall simply assume that the Universe was in an extremely hot and dense state. We shall return to consider these very early times again in Chapter 8 when we look at the way in which theoretical physicists are attempting to develop theories that describe this era.

6.3.4 Inflation and the end of grand unification

$t \sim 10^{-36}$ s, $T \sim 10^{28}$ K, $E \sim 10^{15}$ GeV

When the age of the Universe was about 10^{-36} s, the high-energy conditions under which the strong and electroweak interactions were unified came to an end. After this time, these two types of interaction would become distinct from one another.

Although the typical interaction energies at this time ($\sim 10^{15}$ GeV) are far in excess of laboratory experiments, physicists do have a theoretical framework within which some predictions about this era of cosmic history can be formulated. In fact, it was this approach that, in 1980, resulted in a significant advance in cosmological theory. A theoretical physicist named Alan Guth was tackling a problem which was that as the grand unified era came to an end, it seemed as though a vast number of particles called *magnetic monopoles* should be formed. Such a particle should not decay, and so, if the theory was correct, these magnetic monopoles should be easily detectable in the present-day Universe. The problem was that no such particle had ever been found.

The solution that Guth proposed was quite remarkable. It was based on an analysis of the behaviour of the vacuum. We saw in Chapter 5 that the dark energy may arise from the energy of the vacuum. Although we have no well-developed theory to explain the energy of the vacuum, Guth found that at the end of the grand unified era the vacuum could have had a different, and substantially higher energy density than it does in the present-day Universe. This peculiar state is referred to as the *false vacuum* to distinguish it from the *true vacuum*. This situation would not last for long – in a very short time the energy of the vacuum would drop to the value that it is observed at today – but during that time, something very dramatic would happen.

A high value of the energy density of the vacuum has the same physical effect as a high value of the cosmological constant. (Strictly speaking, the energy density of the false vacuum should not be referred to as being due to a ‘cosmological constant’ since it is hardly constant!) You saw in Chapter 5 that the cosmological model in which the cosmological constant plays a dominant role is the de Sitter model. The evolution of this model is described by Equation 5.11

$$R \propto e^{Ht} \qquad \text{where } H = \sqrt{\frac{\Lambda c^2}{3}} \tag{5.11}$$

In this model the scale factor undergoes exponential expansion. What Guth proposed was that in a very short interval of time – maybe only lasting from $t \sim 10^{-36}$ to 10^{-34} s – the Universe underwent a period of exponential expansion. During this time the scale factor increased by an enormous amount. The theory was not sufficiently well developed to say by exactly how much the scale factor would have increased, but such a process could have caused the scale factor to increase by a factor of as much as 10^{50} (Figure 6.9). This dramatic episode of expansion is termed **inflation**.

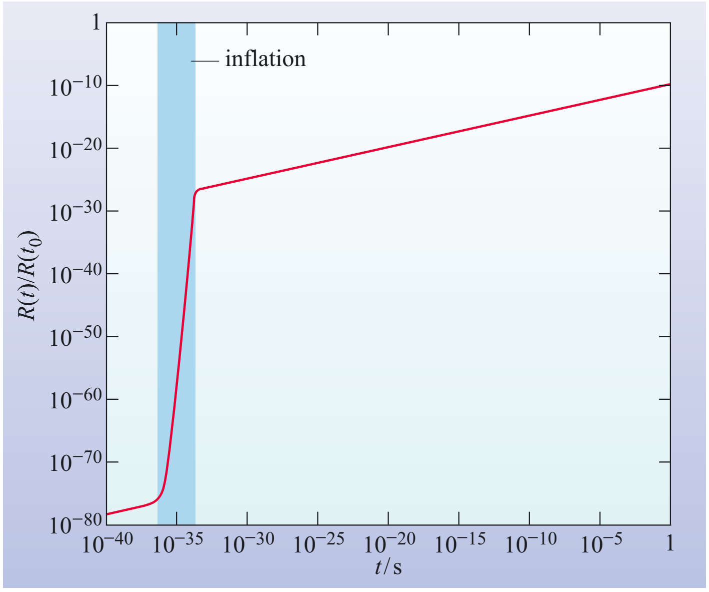


Figure 6.9 The evolution of the scale factor with time including the process of inflation. Note that the numerical values shown here are highly speculative.

So how would inflation solve the monopole problem? Well, the rapid expansion of space during inflation would result in any particles being swept apart from one another. Even if monopoles were abundant prior to inflation, the rapid expansion of space would spread them out so much that there would be a negligible chance of our detecting one in the present-day Universe.

The inflationary scenario has an important consequence for the material content of the Universe. As the period of inflation came to an end, the energy of the vacuum had to drop to the level we see today. So, as the vacuum made a transition to its ‘true vacuum’ state, energy was released and formed particle–antiparticle pairs. According to the inflationary model, the vast majority of particles in the Universe were created from the energy released as inflation came to an end. Thus, the matter now in the Universe may be a product of inflation.

It should be stressed that Guth’s original formulation of the inflationary model should not be considered to be a complete and consistent theory. Rather, it should be viewed as a starting point for exploring a new paradigm in cosmology. In fact, the exact mechanism behind inflation is essentially unknown. In view of this it might seem odd that such prominence is given to the inflationary model until it is appreciated that the *effects* of inflation – regardless of the underlying mechanism – provide solutions to a series of cosmological problems. We shall discuss these problems, and how inflation resolves them, in Chapter 8, but for now we shall simply assume that the process of inflation did occur.

After the process of inflation, the energy released by the false vacuum would have eventually formed all types of quarks and leptons (including their antiparticles). Furthermore, the numbers of quarks and antiquarks would have been almost, but not quite, equal, and a similar condition would have held true for leptons and antileptons. This slight inequality between matter and antimatter in the Universe is thought to have originated in physical processes that occurred at this time. (More will be said about this in Chapter 8.) As we shall shortly see, if this imbalance had not existed, then the present day cosmos would have contained no baryonic matter, and we would not be here to speculate on the origin of the Universe!

Thus, the time at which grand unification came to an end is suspected of playing a major role in the evolution of the Universe. However, from this time at around $t \sim 10^{-34}$ s to the time that the electromagnetic and weak interactions became distinct at $t \sim 10^{-12}$ s, no new physical processes occurred. This interval is often referred to as *the desert*.

6.3.5 The end of electroweak unification

$t \sim 10^{-12}$ s, $T \sim 10^{16}$ K, $E \sim 10^{3}$ GeV

The desert came to an end as electromagnetic and weak interactions became distinct. In contrast to the end of the grand unification, it is not thought that this caused any effects akin to inflation. The constituents of the Universe immediately after this time would have continued to be all types of quark and lepton and their antiparticles. There would also have been photons, and particles that mediate the strong interaction between quarks. However, the temperature was now too low for the creation of W^+, W^- and Z^0 bosons, so the particles that mediate the weak interaction would essentially disappear, thus separating the weak and electromagnetic interactions.

6.3.6 The quark–hadron transition

$t \sim 10^{-5}$ s, $T \sim 10^{12}$ K, $E \sim 1$ GeV

In the present-day Universe, quarks are never seen in isolation – they are always confined within particles called hadrons (see Box 6.2). In the high energy conditions of the early Universe however, quarks were not bound into hadrons; they existed as free individual particles. The existence of free quarks and antiquarks came to an end when the Universe cooled to such an extent that the typical interaction energy was about 200 MeV. At this stage the Universe underwent a phase transition (a process akin to the freezing of water to form ice) in which the quarks became bound into hadrons. This particular phase transition is called the **quark–hadron phase transition**. Although many different types of hadron were formed in this process, there are only two types of hadron that are stable enough to have any long-lasting effect on the composition of the Universe; the proton and the neutron.

The proton (and its antiparticle – the antiproton $\overline{\mathrm{p}}$) is, as far as is known, a stable particle. Both the proton and antiproton can participate in reactions with other particles but, left to themselves, no proton or antiproton has ever been observed to spontaneously decay. The fact that hydrogen exists in copious amounts in the present-day Universe is testament to the stability of the proton. In fact, there is some belief that the proton may decay on very long timescales – but experimental searches for proton decay have shown that its half-life must exceed 10^{33} years. The current age of the Universe is about 1.4×10^{10} years, so as far as this discussion is concerned we can assume the proton to be a stable particle.

Unlike the proton, the neutron (and its antiparticle – the antineutron $\overline{\mathrm{n}}$) is unstable: an isolated neutron will undergo the β^--decay reaction

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e} \tag{6.23}$$

However, the half-life of the neutron is 615 s, and this is a very long time in comparison to the timescale on which the Universe is changing (remember that we are discussing processes that occur within about 10^{-4} s). So to a good approximation, the effect of this decay process can be ignored at this time – and we can say that the neutron is relatively stable.

The mass energies of the proton and the neutron are 938 and 940 MeV respectively. At the time that free quarks became bound into hadrons, the typical interaction energy was too low for proton–antiproton pairs to be produced. Thus, protons and antiprotons would have disappeared from the Universe as they annihilated one another according to the reaction

$$\mathrm{p} + \overline{\mathrm{p}} \rightarrow \gamma + \gamma \tag{6.24}$$

while there would have been no significant counter-conversion of photons into proton–antiproton pairs.

A reaction similar to that shown in Equation 6.24 would also have occurred for neutrons and antineutrons. Prior to this time, baryonic matter was in the form of particles and their antiparticles (either quarks or baryons), but at around this time the majority of such particles annihilated one another. Now we can appreciate the significance of the slight imbalance between matter and antimatter that had been present since the grand unified era. If there had been no imbalance, then at this time, all of the protons would have annihilated with an equal number of antiprotons, and a similar annihilation process would have resulted in the disappearance of all neutrons and antineutrons. The result would have been a Universe that contained no baryonic matter. Within such a Universe there would, of course, be no galaxies, stars, planets or life, so the fact that there was such an imbalance between matter and antimatter was vital to the Universe ending up as we observe it today.

The magnitude of this imbalance between the number of particles and antiparticles is small, but it can be measured from present-day observations. Neither the number of baryons nor the number of photons in a co-moving volume has changed significantly since this time. The present-day ratio of the number of CMB photons to the number of baryons thus provides an estimate of the imbalance between baryons and antibaryons at this time.

At present, there are approximately 10^9 photons in the cosmic microwave background for every stable baryon (proton or neutron) in the Universe.

Thus for every 10^9 baryon–antibaryon annihilation reactions that occurred there would have been one proton or neutron left over.

We have seen that quarks became confined into hadrons and that these hadrons decayed or annihilated one another leaving a residual number of relatively stable protons and neutrons. However, these protons and neutrons were not inert – they could undergo the following reactions that transformed one into the other.

$$\overline{\nu}_\mathrm{e} + \mathrm{p} \rightleftharpoons \mathrm{n} + \mathrm{e}^+ \tag{6.25a}$$

$$\nu_\mathrm{e} + \mathrm{n} \rightleftharpoons \mathrm{e}^- + \mathrm{p} \tag{6.25b}$$

When the age of the Universe was 10^{-2} s, there were large numbers of neutrinos, antineutrinos, electrons and positrons available for such reactions, and the rate of these reactions was high. Consequently the temperature of the neutrinos (as defined by the distribution of their energies) would have been the same as the temperature of the baryonic matter and the temperature of the radiation (as defined by the black-body spectrum).

In the following section (Section 6.4) we shall consider situations in which protons and neutrons participate in fusion reactions. The outcome of these reactions depends on the ratio of the number density of neutrons to the number density of protons (n_n/n_p), and so it is of interest to follow how this ratio varies with time. While the reactions described by Equation 6.25 were occurring, the ratio n_n/n_p depended on the difference in mass energy between these two types of baryon. The proton has a rest mass energy of 938.27 MeV whereas the neutron has a rest mass that is 1.29 MeV greater than this. When the interaction energy was much greater than this difference, i.e. much more than about 1 MeV, then the number densities of protons and neutrons would have been equal. However, once the interaction energy became similar to this energy difference, the number density of neutrons fell below that of protons. At $t = 10^{-2}$ s, when the interaction energy was 10 MeV, the value of $(n_n/n_p) \approx 0.9$, but by $t = 0.1$ s, the typical interaction energy had fallen to 3 MeV and the neutron to proton ratio was $(n_n/n_p) \approx 0.65$.

6.3.7 Neutrino decoupling and electron–positron annihilation

$t \sim 1$ s, $T \sim 1.5 \times 10^{10}$ K, $E \sim 1$ MeV

By the time that the Universe reached an age of 0.7 s, conditions had changed to such an extent that some of the reactions described in Equation 6.25 no longer occurred. In particular, the probability of a neutrino (or antineutrino) interacting with another particle dropped as the density of the Universe decreased. Consequently the reactions shown in Equations 6.25a and 6.25b would only operate from right to left. This was the last occasion on which the bulk of the neutrinos in the Universe underwent any interaction apart from being influenced by gravitational fields. The effective end of the interaction between neutrinos and other particles is termed **neutrino decoupling**. As a result of this, huge numbers of neutrinos, usually referred to as *cosmic neutrinos* started to travel unimpeded through the Universe. They are thought to have been doing so ever since.

Just after neutrino decoupling, when the age of the Universe was about 1 second, the falling temperature of the Universe corresponded to a mean interaction energy of about 1 MeV, which is the energy required for the formation of an electron–positron pair (see Question 6.6). As the temperature fell further, electrons and positrons began to disappear because no new e^+e^- pairs were being created, whereas the annihilation reaction

$$e^+ + e^- \rightarrow \gamma + \gamma \qquad (6.21b)$$

was continuing. The number of electrons and positrons decreased in a dramatic fashion. As was the case when baryons annihilated, there was a slight excess of matter over antimatter – a surplus of one electron for every 10^9 or so annihilation events. The number of negatively charged electrons that were left over is believed to be exactly the number to balance the charge of all the positively charged protons that were left over earlier, thus making the matter in the Universe electrically neutral overall.

An important effect of electron–positron annihilation is that energy was released and this would have been rapidly shared-out amongst the photons, baryons and remaining electrons. Because of this release of energy there was a short interval in which the temperature did not decrease as rapidly as Equation 6.19 would predict, and this is one reason why that equation was described as being approximate.

The process of electron–positron annihilation also leads to a prediction about cosmic neutrinos.

- ■ Would the energy that is released by electron–positron annihilation be transferred to the neutrinos?
- ❑ No. We have just seen that neutrinos effectively stop interacting with the other constituents of the Universe just before electron–positron annihilation occurs.

So cosmic neutrinos do not gain any energy from the process of electron–positron annihilation. Consequently the temperature of cosmic neutrinos should be slightly lower than that of the background radiation. It is predicted that at the present time the cosmological background of neutrinos should have a temperature of about 1.95 K. Experimental confirmation of this would provide strong evidence that the big bang scenario that is described here is correct, but unfortunately, the detection of such low-energy neutrinos is unfeasible at present.

The disappearance of all of the positrons and most of the electrons further restricted the reactions shown in Equation 6.25 that converted protons into neutrons and vice versa. At the time that these reactions stopped completely, the ratio of the number density of neutrons to the number density of protons had a value of $(n_n/n_p) \approx 0.22$, i.e. for every 100 protons in the Universe there were 22 neutrons.

- ■ There was however one reaction that causes neutrons to transform into protons which did not stop. Which reaction was this, and why didn't it stop?
- ❑ The reaction that continues is the β^--decay of the free neutron (Equation 6.23). It did not stop because, unlike the reactions in Equation 6.25 it does not require any other reactant apart from the neutron itself.

Thus, starting from a value of $(n_n/n_p) \approx 0.22$, the number of neutrons started to drop. Unless some new process intervened, the neutrons would have all decayed and we would have a Universe in which the only element that could form would be hydrogen. The way in which this fate was avoided is the next part of our story.

6.4 Nucleosynthesis and the abundance of light elements

$$t < \text{a few hundred seconds},\ T > \text{a few} \times 10^8\,\text{K},\ E > \text{a few} \times 10^4\,\text{eV}$$

We have already seen that conditions in the early Universe led to a situation, such that at $t \sim 1$ s, the temperature was about 10^{10} K and the baryonic matter in the Universe was in the form of protons and neutrons. At this time the physical conditions became suitable for the onset of nuclear fusion reactions which lead to the formation of nuclides with a higher atomic mass than hydrogen. Such a process is believed to have occurred and is called **primordial nucleosynthesis** – a term that distinguishes it from the processes of stellar nucleosynthesis that create elements within stars.

There are some distinct differences between the nucleosynthetic processes that could have occurred in the early Universe and those which occur within stars.

One difference is that the conditions in the Universe were changing rapidly, as the following question illustrates.

QUESTION 6.7

Find the time t at which the temperature was (a) 10^9 K, and (b) 5×10^8 K.

As the answer to Question 6.7 shows, the temperature of the Universe dropped markedly in the first few hundred seconds after $t = 0$. In order for nuclear fusion reactions to have had a significant effect they must have progressed at a rapid rate, and this would have required temperatures in excess of 5×10^8 K. This is in marked contrast to the conditions in the cores of stars where fusion reactions progress at a relatively leisurely rate in lower temperature conditions.

As time progressed in the early Universe, one nuclear reaction that did not require high temperatures, the β^--decay of free neutrons, was proceeding. However, the presence of a large number of free neutrons highlights another difference between the early Universe and stellar cores – that of composition. As we shall now see, it is the declining number of free neutrons that plays an important role in determining how many nuclei can be formed before fusion reactions become ineffective at a temperature of about 5×10^8 K.

6.4.1 The formation and survival of deuterium

The first fusion reaction that could occur was that between a proton and a neutron to form a nucleus of deuterium (which is referred to as a **deuteron**). This is the neutron capture reaction:

$$\mathrm{p} + \mathrm{n} \rightleftharpoons {}^{2}_{1}\mathrm{H} + \gamma \tag{6.26}$$

Note that this is a reversible reaction: the deuteron can be broken apart by γ-rays in a process called **photodisintegration**. In order to cause the photodisintegration of a deuteron, an incident photon must have an energy that exceeds 2.23 MeV. Although at $t = 1$ s the average interaction energy is less than this, there were so many photons in comparison to the number of baryons, that there was a sufficient number of photons with energies greater than 2.23 MeV (i.e. well above the average value) to rapidly destroy any deuterium that formed. However, as the Universe continued to expand and cool, the average photon energy decreased. This decrease allowed deuterium to survive from about $t = 3$ minutes onwards.

To investigate this in more detail we need to know, for a given volume and at a given time (or temperature), what fraction of the photons have sufficiently high energy to cause the photodisintegration of deuterium. A quantity called the **photon energy distribution function** tells us this; it is defined as the fraction of the total number of protons that lie within a narrow energy range, divided by the width of that range. This definition may sound similar to the definition of the spectral flux density (F_λ) – and indeed there is a straightforward mathematical relationship between the two. As in the case of the spectral flux density, the photon energy distribution of a black-body source is a smooth function that has a peak value that depends on temperature.

Figure 6.10a shows the photon energy distribution expected over a range of temperatures in the early Universe. In all cases, of course, the photon energies follow a black-body distribution.

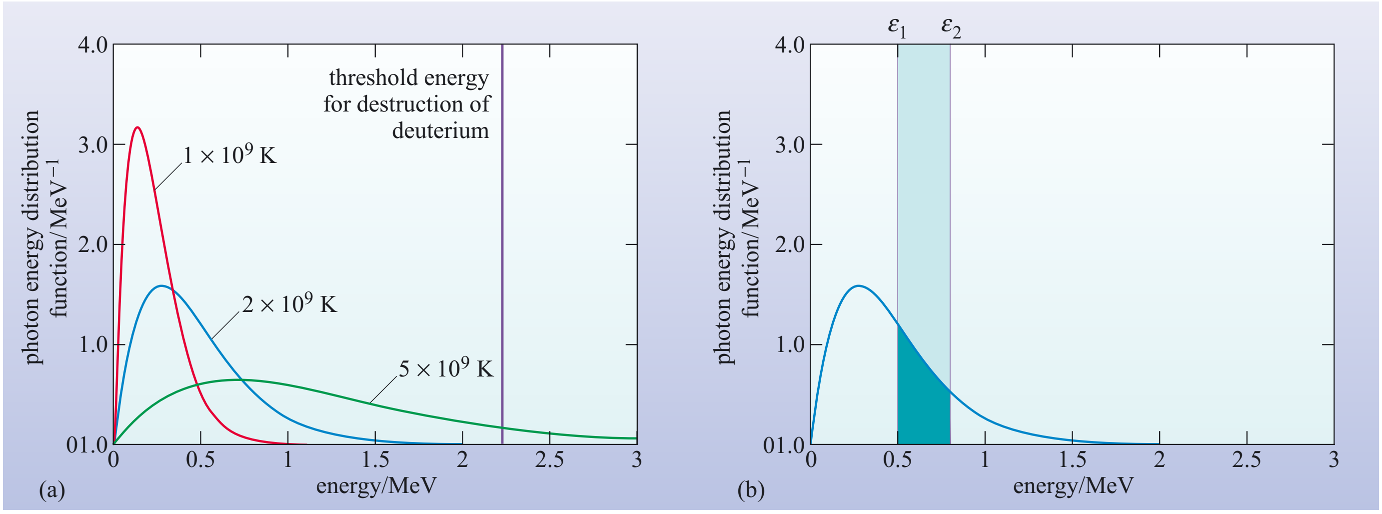


Figure 6.10 (a) Photon energy distributions at various temperatures. (b) The area under the curve of the photon energy distribution and between two energies, indicates the fraction of photons with energies in that range.

- By inspection of Figure 6.10a state qualitatively how the energy at peak of the photon energy distribution of a black-body source varies with temperature.

- The peak energy of the photon energy distribution of a black-body source decreases as the temperature decreases.

The key feature to note about the energy distribution function is that it indicates the fraction of photons that have energies in a certain range. The fraction of photons with an energy between ε_1 and ε_2 is given by the area under the curve between these two limits as shown by the shaded area in Figure 6.10b. Note that the units of the photon energy distribution function are 'per unit energy interval', i.e. $(\text{MeV})^{-1}$. Thus the area under the curve has units which are given by $(\text{MeV}) \times (\text{MeV})^{-1}$: it has no units, as would be expected for a quantity that represents the fraction of photons.

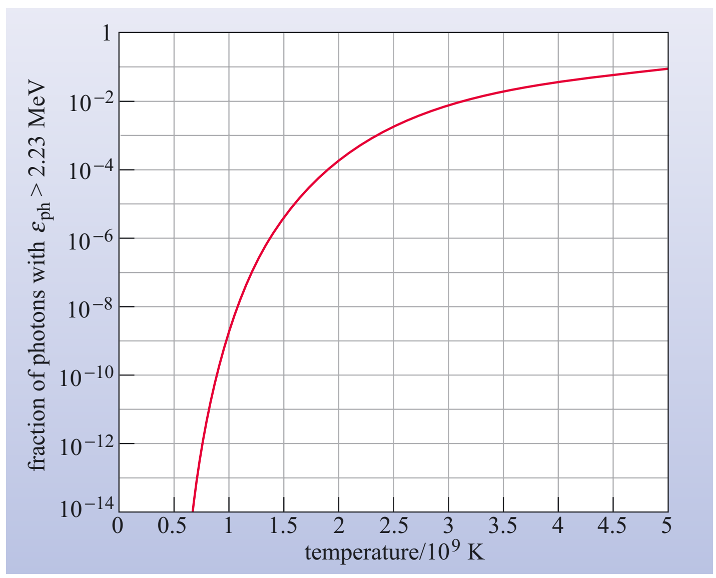


Figure 6.11 The variation with temperature of the fraction of photons with energies greater than 2.23 MeV in a black-body distribution of photon energies.

Returning now to the question of the destruction of deuterium, we know that the ratio of photons to baryons is about 10^9. Thus, if the photon energy distribution were such that more than about 1 in 10^9 photons had an energy greater than 2.23 MeV, then there would have been a sufficient number of energetic photons to destroy any deuterons that may have formed.

Conversely, we can say that deuterons only began to survive once the temperature was such that the fraction of photons with energies greater than 2.23 MeV became less than 10^{-9}. Figure 6.11 shows how this fraction varies with temperature.

■ From Figure 6.11 estimate the lowest temperature at which deuterium can survive, i.e. at which it does not undergo photodisintegration.

❑ From Figure 6.11 the temperature at which this fraction is 10^{-9} is 1.0×10^9 K.

Thus significant quantities of deuterium started to build up only after the temperature dropped below about 1.0×10^9 K. This temperature was reached when the age of the Universe was about 200 s (see answer to Question 6.7(a)).

It is interesting to compare the mean photon energy of a the black-body distribution with the energy required for the photodisintegration of deuterons. For a black-body distribution of photons, the mean photon energy $\varepsilon_{\text{mean}}$ is related to the absolute temperature T by the relation

$$\varepsilon_{\text{mean}} = 2.7kT \tag{6.27}$$

Where k is the Boltzmann constant. So at the time that deuterium starts to build up, (when $T = 1.0 \times 10^9$ K) the mean photon energy is $\varepsilon_{\text{mean}} = 0.233$ MeV. The ratio between the energy that can cause the photodisintegration of a deuteron and the mean photon energy is therefore 2.23 MeV/0.233 MeV = 9.6. Thus, the process of photodisintegration did not stop until the mean photon energy was about a factor of ten lower than the photodisintegration energy. This highlights the fact that because there are vastly more photons than baryons, the very small fraction of photons that have energies much higher than the mean photon energy can have a significant effect on physical processes.

The survival of deuterium has been considered in some detail, since similar considerations about the relative numbers of energetic photons play an important role at a much later time in the history of the Universe – the epoch at which electrons and ions combined to form neutral atoms.

6.4.2 Primordial nuclear reactions

As soon as there was a significant build up in the abundance of deuterium, other nuclear reactions could then proceed. In particular, there were several series of reactions that form the very stable nuclide helium-4 (i.e. ${}^4_2\text{He}$).

For instance, an isotope of hydrogen called tritium (${}^3_1\text{H}$) was formed by

The significance of the colour-coding is explained below.

$$\begin{aligned} &{}^2_1\text{H} + \text{n} \rightarrow {}^3_1\text{H} + \gamma \\ &{}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_1\text{H} + \text{p} \end{aligned} \tag{6.28a}$$

and the tritium thus formed could then undergo reactions to produce helium-4 as follows

$$\begin{aligned} &{}^3_1\text{H} + {}^2_1\text{H} \rightarrow {}^4_2\text{He} + \text{n} \\ &{}^3_1\text{H} + \text{p} \rightarrow {}^4_2\text{He} + \gamma \end{aligned} \tag{6.28b}$$

However deuterium also reacted to produce helium-3

$$\begin{aligned} &{}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + \text{n} \\ &{}^2_1\text{H} + \text{p} \rightarrow {}^3_2\text{He} + \gamma \end{aligned} \tag{6.28c}$$

and this isotope of helium could undergo reactions to form helium-4 as follows

$$ {}^{3}_{2}\mathrm{He} + \mathrm{n} \rightarrow {}^{4}_{2}\mathrm{He} + \gamma \tag{6.28d}$$

$$ {}^{3}_{2}\mathrm{He} + {}^{2}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He} + \mathrm{p}$$

These were the dominant reactions that led to the production of helium-4. (Although the direct fusion of two deuterons to form helium-4 was also a possible reaction, this did not occur to any great extent.)

If the plethora of reactions bothers you, you may be relieved to note that there are only four types of reaction at work here. The reaction equations have been colour coded to illustrate this. The reactions are of the following types:

- A neutron is captured and a photon is emitted (colour-coded green).
- A proton is captured and a photon is emitted (colour-coded red).
- A deuteron is captured and neutron is emitted (colour-coded blue).
- A deuteron is captured and proton is emitted (colour-coded purple).

The major product of primordial nucleosynthesis was helium-4. The fact that nucleosynthesis did not progress to produce large quantities of nuclides with higher mass numbers is due to two factors. Firstly, the rate at which two nuclei will fuse together depends very strongly on the temperature, and higher temperatures are required to fuse nuclei of higher atomic number. Because the deuteron is easily photodisintegrated, the process of nucleosynthesis could only start once the temperature was relatively low. As a consequence, the rate of fusion reactions that involved nuclides other than hydrogen and helium would have been very low.

A second factor is the lack of any stable nuclide with mass number 5 or 8. The lack of a stable nuclide with mass number 5 means that helium-4 could not react with the two most abundant species – protons and neutrons. This hurdle could, however, be overcome by reactions that involve tritium or helium-3,

$$ {}^{4}_{2}\mathrm{He} + {}^{3}_{1}\mathrm{H} \rightarrow {}^{7}_{3}\mathrm{Li} + \gamma \tag{6.29}$$

$$ {}^{4}_{2}\mathrm{He} + {}^{3}_{2}\mathrm{He} \rightarrow {}^{7}_{4}\mathrm{Be}\ \text{(unstable)}$$

$$ {}^{7}_{4}\mathrm{Be} + \mathrm{e}^{-} \rightarrow {}^{7}_{3}\mathrm{Li} + \nu_{\mathrm{e}}$$

One further reaction that is worth noting is that lithium-7 can react with a proton, but the result is the destruction of the newly formed lithium and the formation of two nuclei of helium-4

$$ {}^{7}_{3}\mathrm{Li} + \mathrm{p} \rightarrow {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} \tag{6.30}$$

The yield of lithium was small. The relative amount of lithium formed can be quantified by the mass fraction, i.e. the proportion of the baryonic mass that was in the form of this element. Since lithium is the only element heavier than hydrogen or helium that is formed at this time, the mass fraction of lithium corresponds to the metallicity Z. Primordial nucleosynthesis created lithium such that the metallicity was less than 10^{-9}, but as we will see, the abundance of lithium provides a useful way of probing the conditions during the first few minutes of the big bang.

It should be noted that the fusion reactions that are outlined above only operated for a brief period of time. By the time the Universe reached an age of about 1000 s (i.e. about 17 minutes) and had a temperature of about 5×10^8 K, all such reactions effectively ceased. However, this brief spell of history left a signature, in terms of the abundances of light elements, that can be read today.

6.4.3 The primordial abundance of helium

We can now investigate how much helium would have been produced in the first few minutes of the big bang. The approach that we take here is to estimate the mass fraction of helium (Y) that would have been produced given the processes outlined above.

The starting point for the calculation is the ratio of the number density of neutrons to the number density of protons.

- ■ What was the value of this ratio at the time that the reactions described by Equation 6.25 came to a halt. At what time did this happen?
- ❑ It was stated above that $n_n/n_p \approx 0.22$ at the time that these reactions stopped. This occurred a time $t \approx 0.7$ s.

The temperature at this time was much higher than the maximum temperature at which deuterium can survive.

- ■ Why is the temperature at which deuterons can survive relevant to the production of helium?
- ❑ The formation of deuterons is the first step in the sequence of nucleosynthesis reactions that leads to helium.

So there was a delay before helium synthesis could start. During this time the ratio n_n/n_p does not remain constant because the isolated neutron is not a stable particle. Free neutrons undergo β^--decay with a half-life of 615 s. The next step is to calculate how much time elapses from the instant at which $n_n/n_p \approx 0.22$ and the time at which deuterium can form.

- ■ At what time does the temperature drop to the point at which deuterium can survive? (*Hint*: See Question 6.7.)
- ❑ Deuterium survives once the temperature drops below 1×10^9 K. This temperature is reached at $t = 225$ s. (This was the answer to Question 6.7(a).)

Thus there was an interval of a few hundred seconds before helium production could start in earnest. The way in which the neutron number dropped with time is illustrated in Figure 6.12, which shows the fraction of the original sample of neutrons that would have remained at a given time.

QUESTION 6.8

Calculate the value of n_n/n_p at the time when deuterium started to be formed in significant amounts. (Figure 6.12 provides a way of estimating the number of neutrons that decay in a given time. Take care in calculating n_n/n_p that you account for particles that are created as well as those that disappear!)

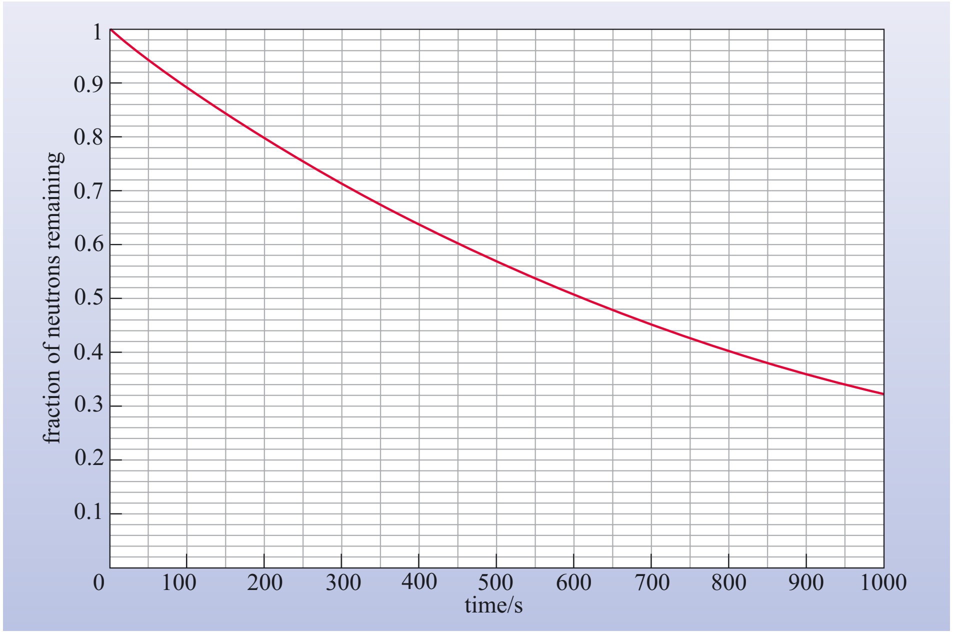


Figure 6.12 The fraction of the number of neutrons that would remain in a sample after a given interval of time. (For use with Question 6.8.)

The result of Question 6.8 shows that when deuterium began to be formed, the ratio of the neutron and proton number densities, n_n/n_p, had a value of about 0.16. The final stage of the calculation is to work out the mass fraction in helium that arises from this value. You have already seen that the major product of primordial nucleosynthesis was helium-4, and as far as calculating the helium abundance is concerned, it is a reasonable approximation to assume that *all* of the neutrons that were present ended up in nuclei of helium-4.

The quantity that we wish to calculate is the mass fraction contained in helium-4

$$Y = \frac{\text{mass of helium in a sample}}{\text{total mass of baryonic matter in the sample}} \tag{6.31}$$

For the purposes of this calculation the sample can be taken to be the baryonic matter in any co-moving volume. There are two contributions to the mass of this sample: that due to hydrogen and that due to helium (which, for simplicity, we assume here to be purely helium-4). If the number of hydrogen and helium nuclei present in our sample, after all nucleosynthesis reactions have stopped, are N_H and N_{He} respectively, then the mass fraction in helium can be written as

$$Y = \frac{N_{He} m_{He}}{N_H m_H + N_{He} m_{He}}$$

Where m_H and m_{He} are the masses of the hydrogen and helium nucleus respectively. This equation can be simplified by making the approximation that $m_{He} \approx 4m_H$.

$$Y = \frac{4N_{He}}{N_H + 4N_{He}} \tag{6.32}$$

Since there are two neutrons in each helium nucleus, the number of helium nuclei is simply half the number of neutrons. The number of hydrogen nuclei is the number of protons minus the number of protons that are locked up in helium nuclei,

$$N_{He} = N_n/2$$

$$N_H = N_p - 2N_{He} = N_p - N_n$$

These expressions for N_H and N_{He} can be substituted into Equation 6.32 to give

$$Y = \frac{2N_n}{N_n + N_p} = 2\left(\frac{1}{1+(N_p/N_n)}\right) \tag{6.33a}$$

The ratio of the number of protons to neutrons (N_p/N_n) in our co-moving sample is the same as the ratio of the *number density* of protons to that of neutrons (n_p/n_n). Hence

$$Y = 2\left(\frac{1}{1+(n_p/n_n)}\right) \tag{6.33b}$$

QUESTION 6.9

Using the value of n_n/n_p that you obtained in Question 6.8, calculate the value of the mass fraction in helium that you would expect from primordial nucleosynthesis.

The answer to Question 6.9 shows a remarkable result: the hot big bang model predicts that the mass fraction of helium-4 should have a value of about 28%. More refined calculations obtain a value that is lower than this – about 24%. This agrees very well with the observation that the mass fraction of helium-4 in low-metallicity interstellar gas seems to be about 24–25%. Until the development of the hot big bang model, the only other mechanism that was a plausible explanation for the production of helium was stellar nucleosynthesis. While it was known that this process does produce helium, it was a mystery how helium could have an almost identical abundance in every location that astronomers measured. The standard hot big bang model provides a more natural explanation for the abundance of helium-4. Such is the success of this outcome of the model, that it is generally interpreted as a key piece of evidence to support the big bang model.

6.4.4 Abundances of light elements as a cosmological probe

In the previous section we saw how the big bang model predicts the formation of helium-4 with an abundance close to that which is observed in the Universe. We have also seen that primordial nucleosynthesis forms other nuclides apart from helium-4.

■ Apart from helium-4, what other stable nuclides are formed by primordial nucleosynthesis?

❑ Deuterium, helium-3 and lithium-7 are the stable nuclides formed by primordial nucleosynthesis. (Note that tritium is unstable with a half-life of about 12 years.)

The abundances are sensitive to the density of matter that is in the form of protons and neutrons, i.e. the density of the baryonic matter in the early Universe. The density of baryonic matter at any time is usually expressed in terms of a baryonic density parameter $\Omega_b(t)$ which is defined as follows

$$\Omega_b(t) = \frac{\text{density of baryonic matter at } t}{\text{critical density at } t} \tag{6.34}$$

Where the critical density at any time $\rho_{crit}(t)$ is given by Equation 5.26

$$\rho_{crit} = \frac{3H^2(t)}{8\pi G} \tag{5.26}$$

The density of baryons in the early Universe can be related to the present-day density of baryonic material, which is expressed using the current value of the baryonic density parameter $\Omega_b(t_0)$. We shall refer to this value as $\Omega_{b,0}$.

The dependence of the abundances of the light elements on the value of $\Omega_{b,0}$ can be determined by detailed calculations: results of such calculations are shown in Figure 6.13. (The abundances also depend on the value of H_0: the calculation shown assumes a value of $H_0 = 72\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$.) The first notable aspect of Figure 6.13 is that the abundance of helium-4 does not vary dramatically with baryon density. Thus, the helium abundance is not very sensitive to the baryon density.

In contrast to helium-4, the abundances of deuterium and helium-3 *are* sensitive to the baryon density. The mass fraction in both of these nuclides decreases substantially as the baryon density increases. The behaviour of the abundance of lithium is somewhat more complex – the curve shows a dip to a minimum value that occurs at a value of $\Omega_{b,0} \approx 0.02$.

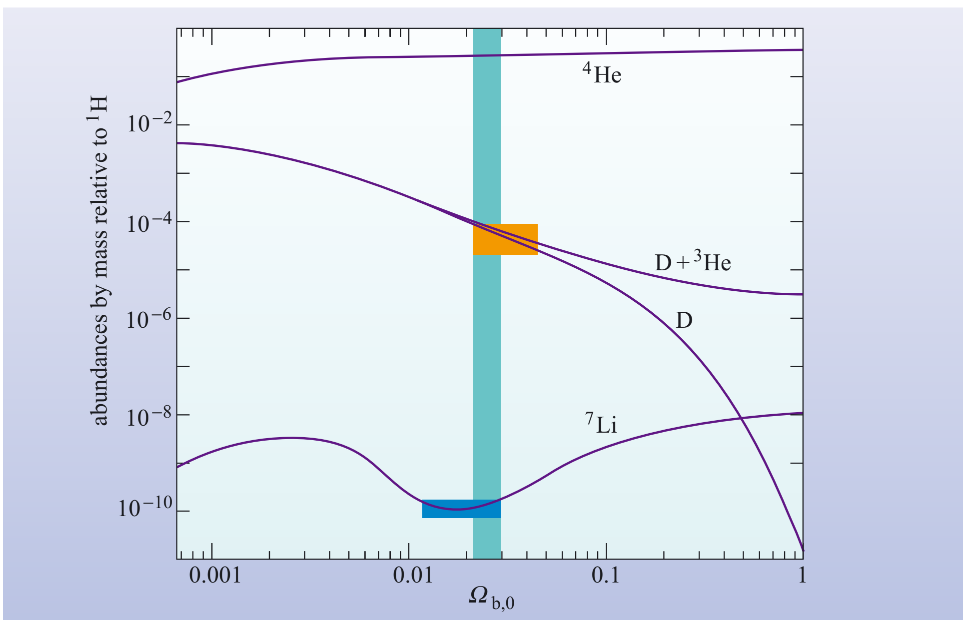


Figure 6.13 The abundances of the light elements as a function of the current value of the baryonic density parameter $\Omega_{b,0}$. Note that D stands for deuterium ($^2_1$H). The smooth curves show the prediction of nucleosynthesis calculations. The orange and blue boxes indicate the ranges of observed abundances. The vertical strip shows the values of $\Omega_{b,0}$ that are consistent with all observations. (Adapted from Walker *et al.*, 1991)

If the abundances of these light elements could be measured in material that has not undergone any change in composition since the time of primordial nucleosynthesis, then we could, in principle, measure the current baryonic density parameter of the Universe. A fundamental problem in this approach is highlighted by the following question.

QUESTION 6.10

Use Figure 6.13 to make an order of magnitude estimate of the maximum metallicity of material that has only undergone primordial nucleosynthesis, and has not been chemically enriched by stars. Is this value of metallicity consistent with the idea that the oldest observed stars were formed from primordial material?

The answer to Question 6.10 shows that even the oldest stars have undergone some chemical enrichment in a previous generation of stars; no stellar material appears to be left over from the big bang that has not undergone some subsequent nuclear processing by stars. Despite this complication, it is possible to make progress in this field by trying to work out how the abundances of deuterium, helium-4 and lithium may have been changed by nuclear processes that might have occurred since the end of primordial nucleosynthesis. When corrections of this sort are made, the result is a range of plausible values for the primordial abundances of deuterium, helium-4 and lithium. These ranges are shown as rectangular regions on Figure 6.13, and the overlap of these regions gives an allowed range of $\Omega_{b,0}$. The result is that $\Omega_{b,0}$ lies in the range 0.02 to 0.03. This immediately indicates that if the Universe has a flat geometry (implying $\Omega = 1$), then at most, 3% of the contribution to the density of the Universe can arise from baryonic matter. Given the difficulties in analysing primordial abundances, these figures should be treated with some caution, but it seems certain that baryonic matter can contribute no more than 5% to the total density of the Universe. As you will see in the next chapter, evidence is growing that $\Omega = 1$, which has the immediate implication that at least 95% of the density of the Universe is in a form that is different to the matter from which stars are made.

QUESTION 6.11

Suppose that the primordial abundances of light elements were measured in three ‘hypothetical universes’ as shown in Table 6.4. What would be the current value of the density parameter $\Omega_{b,0}$ of baryonic matter in each case? Could you determine the value of $\Omega_{b,0}$ from the lithium abundances alone?

Table 6.4 Some abundances in hypothetical universes – for use with Question 6.11.

Hypothetical universe	Lithium abundance by mass relative to hydrogen ($^1_1$H)	Deuterium abundance by mass relative to hydrogen ($^1_1$H)
A	1×10^{-9}	7×10^{-4}
B	1×10^{-9}	1×10^{-5}
C	$< 1 \times 10^{-9}$	1×10^{-4}

6.5 Recombination and the last scattering of photons

$t \approx 3$ to 4×10^5 years, $T \approx 4500$ to $3000\,\mathrm{K}$, $E \sim$ a few eV

After the nucleosynthesis of light nuclei that occurred in the first few hundred seconds after $t = 0$, the Universe expanded and cooled for several hundred thousand years before it underwent another dramatic change. This next big event is called *recombination* and occurred when nuclei and electrons combined to form neutral atoms. In this section we will examine this epoch of the Universe and see how processes that took place at that time account for the cosmic microwave background that is detectable today.

6.5.1 The Universe in the post-nucleosynthesis era

As the temperature of the Universe dropped below $10^8\,\mathrm{K}$, the nuclear reactions that resulted in the formation of light elements came to a halt. The composition of the Universe at this time was: protons; nuclei of deuterium, helium and lithium; electrons; neutrinos; photons; and dark matter particles (whatever they are!). The important interactions were those which shared energy out between the different constituents. In particular, photons and electrons interacted with one another to exchange energy. The electrons also collided with protons and nuclei and this also led to a sharing out of energy. These two types of interaction ensured that the radiation, electrons and nuclei remained in a state of thermal equilibrium.

■ Why were the neutrinos not in thermal equilibrium with the other particles at this stage in the evolution of the Universe?

❑ Because the rate of neutrino interactions was very small. Most neutrinos did not interact with any particle after the time of neutrino decoupling.

As the Universe expanded, the temperature of the background radiation dropped with time. Because the photons greatly outnumbered the electrons and nuclei, the temperature of the electrons and nuclei was kept in step with that of the cooling radiation.

It has already been noted that at early times the expansion of the Universe was dominated by the energy density of radiation. However as you saw in Figure 6.6, the energy density of matter became equal to that of radiation when the scale factor attained about 10^{-4} of its current value, and this would have occurred when the age of the Universe was a few times 10^4 years. Following that event, the rate of expansion would have been dominated by the energy density of matter, and the effect of radiation would have progressively declined. Despite this change in the expansion rate, the temperature would still have been that of the background radiation due to the incessant interactions between photons, electrons and nuclei. This state of affairs would have persisted until the interactions between cosmic background photons and free electrons became negligible.

6.5.2 The era of recombination

An important type of interaction between electrons and nuclei is that which results in the formation of a neutral atom. As an electron becomes bound into an atom, energy is released in the form of a photon. This process is called **recombination**.

As the Universe expanded and cooled, conditions became favourable for recombination to occur. (In this context, the term ‘recombination’ may seem somewhat of a misnomer as electrons and nuclei had never been ‘combined’ as neutral atoms before this time!)

The process of recombination is the opposite of ionization. The reaction can be written as

$$\mathrm{p} + \mathrm{e}^- \underset{\text{ionization}}{\overset{\text{recombination}}{\rightleftarrows}} ({}^1_1\mathrm{H})_{\text{neutral}} + \gamma \qquad (6.35)$$

In the case of neutral hydrogen, the energy required to ionize the atom from its lowest energy (ground) state is 13.6 eV. Thus, if conditions were such that for every atom, there were many photons with an energy of at least 13.6 eV, then the neutral atom would not have survived for long, but would have soon been re-ionized. Alternatively, the atom may have undergone a collision with an electron or proton that could also have supplied the 13.6 eV of energy required to ionize the atom. Either way, at high temperatures, the reaction described in Equation 6.35 favours the production of protons and electrons, and the number density of neutral atoms is exceedingly low.

As the temperature of the Universe fell, the equilibrium of Equation 6.35 shifted to favour the production of neutral atoms and photons. The temperature at which this change occurred was subject to very similar constraints as those you met earlier in connection with the photodisintegration of deuterium. The number of photons exceeds the number of hydrogen atoms by a factor of about 10^9. Thus if only one in 10^9 photons had an energy of 13.6 eV, then this will be sufficient to ionize any neutral hydrogen.

QUESTION 6.12

By analogy with the case of deuterium photodisintegration, estimate the temperature at which recombination of hydrogen occurs.

Recombination started to occur when the temperature was about 4500 K. This temperature was reached when the Universe was about 3×10^5 years old. As neutral atoms formed, the density of free electrons (i.e. electrons that are not bound up in atoms) decreased. This had an important effect on the interrelationship between photons in the background radiation and matter in the Universe. Scattering interactions between free electrons and photons became infrequent. By the time the temperature had dropped to 3000 K, the number density of free electrons was so low that the Universe essentially became transparent and photons could travel unhindered from this time on. As we shall shortly see, the radiation that we observe now as the cosmic microwave background was last scattered at the time of recombination.

The lack of scattering interactions had important thermal and dynamical effects. As we saw above, prior to recombination, the energy of photons was ‘shared’ with the thermal energy of matter in the Universe, and this ensured that matter had the same temperature as the background radiation. After recombination the temperature of the matter and the background radiation evolved independently of one another. The temperature of the background radiation changed with scale factor according to Equation 6.6 – indeed it was from this relationship we were able to infer that the cosmic microwave background that is observed at present implies a hot early Universe.

The important dynamical role of the coupling between photons and electrons relates to the stability of over-dense regions against gravitational collapse. Prior to recombination, the radiation pressure played an important role in opposing the gravitational collapse of over-dense regions. After recombination, the rate of interactions between photons and electrons dropped dramatically, and this suddenly allowed over-dense regions that had been expanding with the Hubble flow to start to begin to contract under gravity. We will explore this aspect of the post-recombination Universe in more detail in Section 6.6, but first we will consider the evolution of the background radiation to form the cosmic microwave background.

6.5.3 The formation of the cosmic microwave background

The major change that occurred to the photons as they travelled after their last scattering was that they were red-shifted by the expansion of the Universe. As has already been mentioned, cosmological expansion causes the wavelengths of photons to increase, and if those photons are distributed according to a black-body spectrum, then this form of the spectrum is retained even though its characteristic temperature is reduced. We have now reached a detailed explanation for the formation of the cosmic microwave background: it is the cooled 'gas' of red-shifted photons that were in thermal equilibrium with matter at the time of last scattering.

QUESTION 6.13

Calculate the redshift at which the last scattering occurred. Assume that at the time of last scattering the temperature of the background radiation was 3.0×10^3 K. (*Hint*: Start by using Equation 6.6 to determine the change in scale factor.)

As the answer to Question 6.13 shows, the last scattering of photons in the cosmic background radiation occurred at a redshift of about 1100.

One way of thinking about how such photons should appear to us now is illustrated schematically in Figure 6.14: the diagram shows our vantage point – 'here and now' – at the centre, and the radial direction from this point represents a direction on the sky. The radial distance from our viewpoint on this diagram represents time, such that the present is at the centre whilst the outermost circle of the diagram is at $t = 0$, i.e. at the first instant of cosmic expansion. Whatever direction we look in we see the radiation from the time that the Universe became transparent. This transition, from being opaque to being transparent occurred when the Universe was about 3×10^5 years old. At times before this, the Universe was opaque due to the scattering of photons by electrons.

Thus the cosmic background radiation appears to come from a spherical surface called the **last-scattering surface**, that is centred on us. Of course, this does *not* mean that we are at the centre of the Universe! An analogy may help here: if you go for a walk on a foggy day when the visibility is 30 m, then you will be able to see up to 30 m away in any direction.

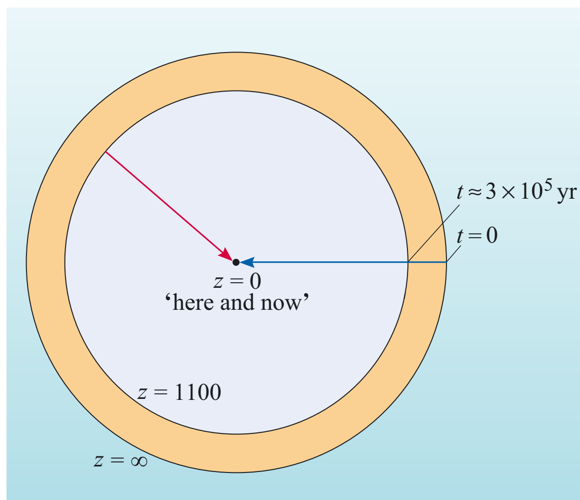


Figure 6.14 A schematic figure showing the origin of photons in the cosmic microwave background as observed from the Earth. In this diagram the radial direction from the Earth represents a direction on the sky and the radial distance from the Earth represents time.

Your view of the 'foggy world' is that you are at its centre, and that it has a radius of 30 m. Of course, this is just *your* view of the world; any other observer in the fog will have a similar view even though they might be somewhere else, see Figure 6.15. Thus the boundary defined by how far any single observer can see has no global significance. Likewise the last-scattering surface simply represents as far as we can see from our location in the Universe. The situation is somewhat different from viewing the world through a uniform fog – since in viewing the CMB we are essentially looking back in time to an era when the Universe was, in a sense, 'foggy'. However, the principle is the same: observers at another location in the Universe would also see a last-scattering surface, but it will not correspond physically to the last-scattering surface that we observe.

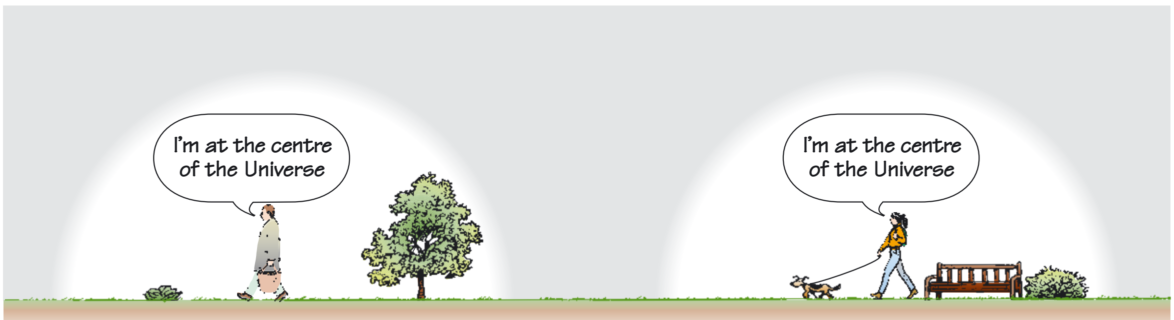


Figure 6.15 Observers walking on a foggy day can see a fixed distance in any direction. Despite the naive claims of the observers, the boundary defined by this distance has no global significance!

6.5.4 Observing the cosmic microwave background

Now that we have developed an explanation for the cosmic microwave background in terms of red-shifted radiation that was last scattered by free electrons at a redshift of $z \approx 1100$, it is appropriate to consider its observed properties in more detail and how such observations are made.

The properties of the cosmic microwave background radiation that have already been described can be summarized as follows:

- The CMB has a black-body spectrum, with a characteristic temperature of 2.725 ± 0.002 K.
- The CMB is highly uniform in every direction that we observe (Section 5.2.3).

We shall shortly see that the CMB is not perfectly uniform, and that this is an important source of information about the Universe. However we shall begin by considering how measurements of the spectrum of the CMB are made.

The cosmic microwave background was discovered in 1965 by Arno Penzias and Robert Wilson (Box 6.3). Penzias and Wilson could not measure the spectrum of this signal since their receiver was tuned to work at a single wavelength (of 7.35 cm), so initially there were no data to support the idea that the background had the spectrum of a black body. However, the surface brightness of the radiation was characteristic of emission from a black body at a temperature of 3 K.

BOX 6.3 THE DISCOVERY OF THE COSMIC MICROWAVE BACKGROUND

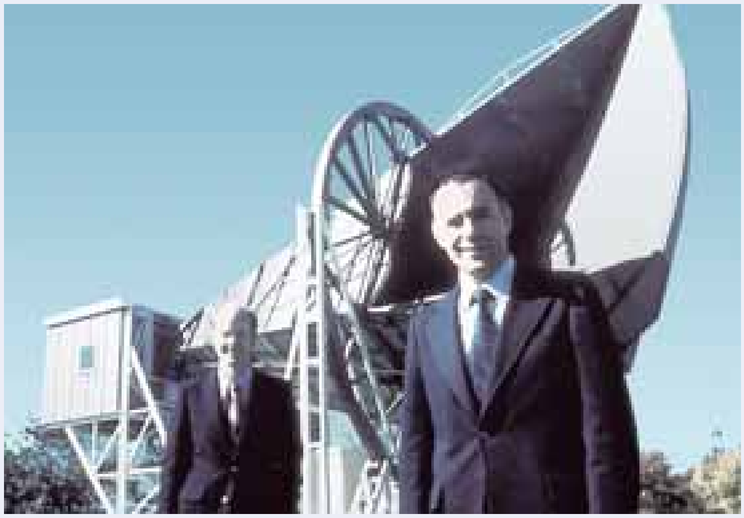


Figure 6.16 Arno Penzias (right) and Robert Wilson (left) in front of the horn antenna with which they discovered the cosmic microwave background. Penzias and Wilson were awarded the 1978 Nobel Prize in Physics for their discovery of the cosmic microwave background. (Bell Laboratories, AT&T)

In 1964, two researchers at Bell Laboratories in Holmdel, New Jersey, Arno Penzias and Robert Wilson (Figure 6.16), were charged with the task of calibrating a radio antenna to be used for telecommunication and galactic radio astronomy. The antenna, which is shown in the background of Figure 6.16, was in the form of a horn: radio waves entered through the aperture and were directed onto a detector at the narrow end of the horn. A feature of this type of design of radio antenna is that the signal that is received at the detector should be relatively free from any contamination that arises from sources that are outside the field of view (this is a problem that most other designs of antenna, such as a dish, do suffer from). Soon after starting, Penzias and Wilson measured the signal from an airborne radio emitter, and found that there was an unexplained source of noise within the system. For the next year or so, they tried to track down the source of this noise; eliminating the possibility that it may be due to faulty electronics, rivets in the antenna – or even pigeon droppings in the horn. One characteristic of this noise was that it seemed to be the same wherever on the sky the antenna was pointed. Another was that the intensity of the signal was as would be expected if the whole sky were a black-body source at a temperature of 3 K.

In seeking an explanation for this signal, Penzias and Wilson were put in touch with Robert Dicke, a professor at Princeton University (which is only a few miles away from Holmdel). Dicke had realized several years earlier that the early stages of the big bang would have been characterized by very high temperatures, and that the radiation from this phase of the history of the Universe should be detectable today with a black-body spectrum and a temperature of a few kelvin. He had also realized that with the then current state of radio technology it should have been possible to detect such a signal. His research group were in the process of building a dedicated detector when Penzias and Wilson contacted him. Although Dicke and his team were beaten to the discovery of the signal, they were in a prime position to offer an interpretation of the result. The discovery was published as two papers in 1965; one by Penzias and Wilson which described only the observational result, and another by Dicke and his co-workers that offered an interpretation of the result as the signal being a relict of the hot big bang.

A further twist in the tale arose after the results were first published. Dicke received a terse letter from George Gamow, a Russian-born American physicist. In 1948, Gamow, with his collaborators, Ralph Alpher and Robert Herman had predicted that the temperature during the early stages of the big bang were sufficiently high that nuclear reactions would have taken place (Section 6.4). Furthermore, he suggested that a remnant of this high temperature phase should exist as background radiation with a temperature of about 5 K. Having made a prediction about the background radiation seventeen years earlier, Gamow was justifiably annoyed that his work had been overlooked. However, in the years since the discovery of the cosmic background radiation, Gamow's contribution to the implications of the hot big bang model have become widely recognized.

As we have already seen, the peak of emission from a black-body spectrum at a temperature of 3 K occurs at wavelengths of about 1 mm. The Earth's atmosphere partially absorbs radiation in this part of the electromagnetic spectrum, which makes it difficult to measure the spectrum of the cosmic microwave background using a ground-based antenna. The obvious solution to this is to place a detector above the Earth's atmosphere. During the 1970s and 1980s measurements were made using balloon-borne instruments that went some way to confirming that the cosmic microwave background has a black-body spectrum. But a major advance came from a satellite-based experiment called the Cosmic Background Explorer or COBE. This mission was launched in 1989, and was designed to measure various characteristics of the cosmic microwave background. One striking result from COBE was the measurement of the spectrum of the microwave background radiation as shown earlier in Figure 6.2. This spectrum follows the theoretical black-body curve to a very high degree – in fact, it is the best example of a black-body spectrum that has ever been observed in nature.

The second major aspect of the cosmic microwave background that has been subject to intense observational scrutiny is its uniformity. The largest variations in the microwave background as we look from one position to another correspond to a variation in temperature of less than one part in 10^3. It is difficult to make absolute measurements of the intensity of the CMB, and so experiments that measure variations in intensity do so by comparing signals from two different regions of the sky. Typical observing strategies involve making a large number of these comparison measurements across the area of sky that is of interest (which may be the entire celestial sphere). In this way, instrumental uncertainties are reduced, and levels of *relative* variation in the CMB can be measured to better than one part in 10^5. We will return to this topic in Chapter 7, but now we will take a first look at the implications of the measured uniformity of the cosmic microwave background.

6.5.5 Interpreting the cosmic microwave background

Much of this chapter has been based on interpreting one aspect of the cosmic microwave background – its spectrum. Thus we have studied the significance of the black-body form of the cosmic background radiation, and explored the consequences of the presence of such radiation in an expanding Universe. However there is a lot more to be learnt from the cosmic microwave background than can be deduced from its spectrum alone. In particular, the uniformity of the cosmic microwave background, and departures from this uniformity, also need to be explained in terms of cosmological processes. It is to this aspect of the cosmic microwave background that we now turn.

The uniformity of the cosmic microwave background

The view of the formation of the cosmic microwave background as illustrated by Figure 6.14 seems reasonable but gives rise to a puzzle. The temperature of the cosmic microwave background is, to a very high degree, uniform in whichever direction we look. The uniformity in the microwave background implies that at the time of recombination, the temperature at any point on the last-scattering surface was also highly uniform. The puzzling aspect of this is that the last-scattering surface is too large to have settled down to a uniform temperature in the 3 to 4×10^5 years that the Universe had been expanding.

A limit to the size of a region that can come to thermal equilibrium can be found by a similar argument to that used in Chapter 3 to constrain the size of AGN. The underlying principle is that a physical signal cannot propagate at a speed that is greater than the speed of light. In this case the 'physical signal' is the heat flow caused by a difference in temperature between two locations. If we consider one location in the Universe at the instant when the expansion of the Universe began, then at subsequent times, the most distant point that could possibly be in thermal equilibrium with our starting point would be at the distance that light could travel in the age of the Universe. This distance, which is the limiting size of a region that can be expected to be in thermal equilibrium at a given time is called the **horizon distance**. This is illustrated schematically in Figure 6.17: the point X represents an initial point and X′ is the same point at recombination (i.e. X and X′ have the same co-moving coordinates). The maximum distance that could be covered by any signal at a given time prior to recombination is shown by the line from X to A, and the distance from X′ to A is the horizon distance at the time of recombination. If the horizon distance is calculated for the last-scattering surface, it turns out that it corresponds to an angle θ on the sky of about 2°. Regions of the last-scattering surface that are separated by more than this angle could not have affected each other.

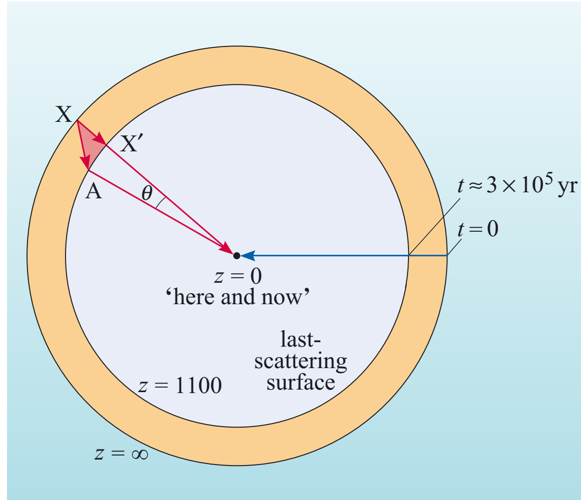


Figure 6.17 The horizon distance at the time of recombination as it appears on the last-scattering surface. Point X represents a location in the Universe just after $t = 0$. X′ is the location with the same co-moving coordinates as X at the time of recombination. The distance from X′ to A is the maximum distance that a signal could travel prior to recombination. The angular separation of points A and X′ is about 2°. (Note that this diagram is not to scale.)

The fact that the microwave background has almost the same temperature over scales that are much greater than the horizon scale of 2° is a problem that the standard big bang model fails to address. We will return to consider a solution to this *horizon problem* in Chapter 8.

Although the cosmic microwave background is remarkably smooth there are departures from perfect uniformity. These variations – or **anisotropies** – are at a very small level. As discussed in the following two subsections, there are two distinctly different mechanisms that give rise to anisotropies in the microwave background.

The dipole anisotropy

Mapping of the cosmic microwave background reveals that in one direction in the sky the temperature is 3.36 mK higher than the mean temperature, whereas in the opposite direction it is 3.36 mK lower than the mean. Between these two directions there is a smooth variation between the maximum and minimum temperatures. This is illustrated by the all-sky map shown in Figure 6.18 – the blue and pink areas are above and below the average temperature respectively. Because the variation is symmetric about two opposite directions, or poles, on the celestial sphere, it is referred to as the **dipole anisotropy**.

Figure 6.18 A map from COBE showing the dipole anisotropy. The elliptical map area covers the entire sky. The blue and pink coloured regions represent temperatures that are higher and lower, respectively, than the mean. (Goddard Space Flight Center)

The interpretation of the dipole anisotropy is that it arises from our motion relative to the co-moving reference frame. Figure 6.19 indicates the effect of observer motion on the measured wavelength of cosmic microwave background photons.

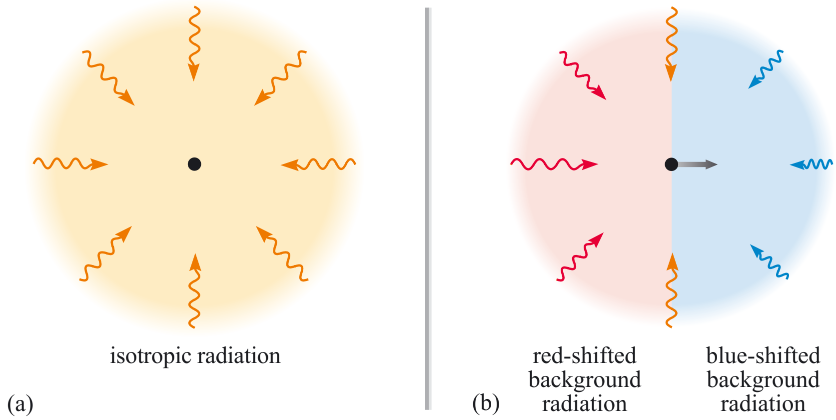


Figure 6.19 The view of the cosmic microwave background according to (a) an observer who is stationary with respect to the frame of reference of co-moving coordinates, and (b) an observer who is moving with respect to this frame of reference. According to the observer in case (a) the background is isotropic, whereas the observer in case (b) measures a blue-shift in the direction of motion and red-shift in the opposite direction.

Figure 6.19 concerns two observers at the same location in space, but who are moving with respect to one another. Observer (a) sees the cosmic microwave background as being perfectly uniform in all directions, whereas observer (b) sees the same last-scattering surface, but the radiation is Doppler shifted. The perfectly uniform case arises for an observer who is stationary in co-moving coordinates, i.e. one who is perfectly following the Hubble flow of the expanding Universe. Any motion relative to this frame of reference, such as that of observer (b), would cause the observer to see the cosmic microwave background blue-shifted in the direction of motion and red-shifted in the opposite direction. When radiation with a black-body spectrum is subject to a Doppler shift, the spectrum retains its black-body form, but the peak wavelength is shifted and the characteristic temperature changes accordingly. Red- and blue-shifts result in lower and higher temperatures respectively. Thus the direction on the sky in which the temperature increases the most corresponds to the direction in which we are moving with respect to the Hubble flow.

■ Why is it unlikely that the Earth would be stationary with respect to the frame of reference of co-moving coordinates? List possible contributions to the Earth's velocity with respect to this frame?

❑ The Earth is unlikely to be stationary in co-moving coordinates because (i) the Earth is in motion around the Sun, (ii) the Sun is in motion around the centre of our Galaxy, and (iii) the Galaxy has a peculiar motion with respect to the Local Group, and (iv) the Local Group is likely to have a random motion with respect to the Hubble flow.

Analysis of the dipole anisotropy shows that the Sun has a speed of about 365 km s^{-1} with respect to the local co-moving frame of reference. When the Sun's motion around the Galaxy, and the Galaxy's motion with respect to the Local Group are accounted for, it is found that the Local Group has a speed of about 630 km s^{-1} with respect to the local co-moving frame of reference.

Intrinsic anisotropies in the cosmic microwave background

Even when the effect of the dipole anisotropy is removed the cosmic microwave background is still not perfectly smooth – it exhibits smaller scale variations in temperature from one point to another. Figure 6.20 shows a map of these fluctuations over a region of sky that is about 10° across. The magnitude of such variations is small: at most they correspond to a fractional change in temperature of only a few parts in 10^5. The significance of these intrinsic anisotropies is that, although they mainly represent temperature variations in the Universe at the time of recombination, they are intimately linked to the small variations in *density* that are believed to have arisen at very early times in the history of the Universe. In particular, it is hypothesized that these density fluctuations may have their origin in the process of inflation that occurred at $t \sim 10^{-36}$ s. We shall consider the origin of these so-called *primordial fluctuations* in Chapters 7 and 8, but we note here that the analysis of intrinsic anisotropies in the cosmic microwave background can potentially provide information about processes in the very early history of the Universe.

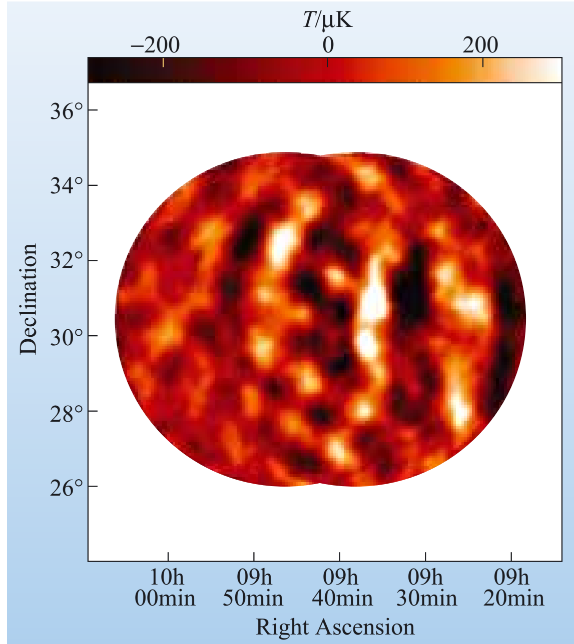


Figure 6.20 A map of the fluctuations in the cosmic microwave background as measured by an experiment called the Very Small Array. (University of Cambridge, Jodrell Bank Observatory, and the Instituto de Astrofísica de Canarias)

Temperature variations in the cosmic microwave background are also useful because the photons that were last scattered at the time of recombination have been moving freely, apart from the effects of gravity, ever since. So the anisotropy pattern can also provide information about the geometry of space–time. For this reason, the analysis of anisotropies in the cosmic microwave background is a key area of observational cosmology, and will be discussed more fully in Chapter 7. However, here we continue our discussion of the evolution of the Universe in the post-recombination era by considering how the small variations in density that were present at recombination helped to form structure in the Universe.

6.6 Gravitational clustering and the development of structure

The small amplitude of temperature anisotropies in the cosmic microwave background indicates that the distribution of baryonic matter in the Universe was very smooth at the time of recombination. This is in marked contrast to the present-day Universe in which the baryonic matter is concentrated into stars, galaxies, clusters and large-scale structure. An outline of how the Universe evolved from an almost, but not quite, uniform state into such structures was given in Chapter 2, and here we consider the formation of structure in the light of our account of the big bang model. Our starting point is to consider a simple, but rather unrealistic scenario in which all the matter in the Universe is assumed to be baryonic. We shall then see how the study of the growth of structure can provide some clues as to the nature of the non-baryonic matter that is actually believed to have been present.

6.6.1 Gravitational collapse in a baryonic matter Universe

You have already seen (Chapter 2) that structure is believed to form as a result of the gravitational collapse of a region that is initially denser than average. The early stages of this process are very gradual. Remember that the Universe is expanding, and initially, over-dense regions will simply expand at a slower rate than average. Furthermore, the density variations in the early Universe are at a low level. Since the density of the Universe is changing with time, it is useful to express the magnitude of density variation using the **relative density fluctuation**, which is defined as follows

For clarity, we have used ρ rather than ρ_m to denote the density of matter in this equation. We shall continue to use this simplified notation in the next chapter.

$$\frac{\Delta\rho}{\rho} = \frac{\text{density within a fluctuation} - \text{mean density of the Universe}}{\text{mean density of the Universe}}$$

The level of the primordial fluctuations suggests that in the early Universe $\Delta\rho/\rho \sim 10^{-5}$.

Whether a particular density enhancement will grow by gravitational collapse depends on the balance between two effects. One effect is the self gravity of the matter within the over-dense region which, of course, tends to cause collapse. The opposing effect is due to pressure which acts to stabilize over-dense regions against collapse. Which of these two effects is dominant under specified conditions depends on the mass of the region. The British scientist Sir James Jeans, who first analysed gravitational collapse in relation to the formation of stars, discovered that a key parameter is a quantity that is now known as the **Jeans mass**. The Jeans mass represents the boundary between two different types of behaviour. If the mass of an over-dense region exceeds the Jeans mass then it will collapse. However, a region that contains less than the Jeans mass would be supported by its internal pressure and hence be stable.

The horizon distance plays an important role in the discussion of stability against collapse. An over-dense region that is larger than the horizon distance cannot be supported by its internal pressure. This is because any changes in pressure are propagated at a speed that is lower than the speed of light. Nevertheless, the relative density fluctuation within this over-dense region *does* grow with time, albeit slowly, since the over-dense region is expanding at a lower rate than the Universe around it. However, the horizon distance increases with time, so eventually, the over-dense region will lie within the horizon distance and can respond to internal pressure changes. After this time, an over-dense region can be stable against collapse provided that the mass within this region is less than the Jeans mass.

The Jeans mass depends on the density and pressure of the region under consideration. In an expanding Universe, the density and pressure change as the scale factor changes. This causes the Jeans mass also to vary with time – this behaviour is shown in Figure 6.21. Note that the Jeans mass increases steadily with scale factor during the radiation-dominated phase, but it then flattens out and then drops at the time of recombination. To help in interpreting this diagram, it is useful also to indicate the criterion that the diameter of an over-dense region must be less than the horizon distance before it can become stable against collapse. To do this, Figure 6.21 also shows a quantity called the **horizon mass** which is the mass contained within a sphere with radius equal to the horizon distance. As expected, the horizon mass increases with time because the horizon distance increases with time (because light has had the time to travel further). The region above the horizon mass line represents over-dense regions that have a diameter greater than the horizon, and as mentioned above, these grow in a rather sedate fashion.

Figure 6.21 implies that, prior to the time when matter and radiation energy densities became equal, any over-dense region with a mass of up to about $10^{16}M_{\odot}$ will at some time be overtaken by the horizon – that is, the mass within the region will be overtaken by the horizon mass. However the horizon mass and the Jeans mass are very close to one another, so almost immediately, the mass of such an over-dense region will also be overtaken by the Jeans mass. Thus, the over-dense region will become stable against collapse, and the amplification of fluctuations as a result of gravitational contraction comes to a halt – at least temporarily.

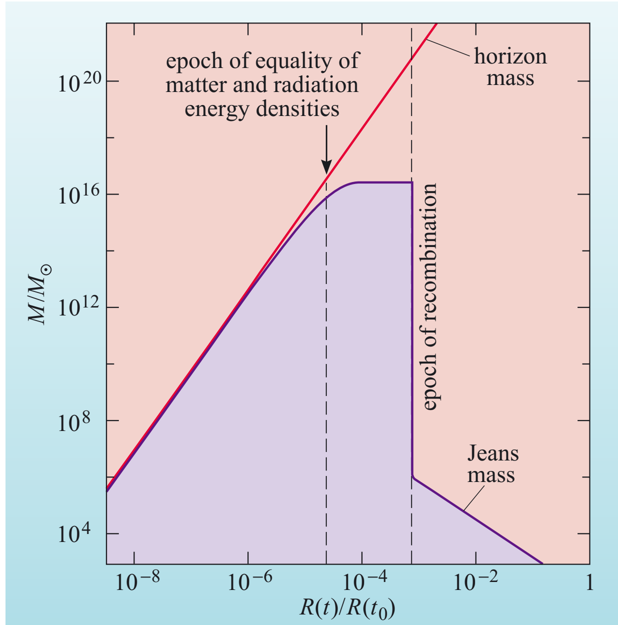


Figure 6.21 The variation of the Jeans mass (purple line) and the horizon mass (red line) as a function of the scale factor ($R(t)/R(t_0)$). Gravitational collapse only occurs for regions of the diagram that are shaded in pink. Regions that are shaded in lilac are stable against collapse. (Adapted from Longair, 1998)

An important change occurs at the time of recombination: the Jeans mass drops dramatically. Prior to recombination the major contribution to the pressure arose from the interaction of photons with free electrons. After recombination electrons become bound into neutral atoms which do not interact with the background photons. The radiation pressure becomes negligible and the only source of support against collapse comes from the thermal pressure of the gas. This thermal pressure is much smaller than the radiation pressure prior to recombination, so much smaller masses suddenly become unstable to gravitational collapse. The Jeans mass drops from a value of about $10^{16}M_{\odot}$ just before recombination to about $10^{5}M_{\odot}$ just after. Thus, any fluctuations in the mass range from $10^{5}M_{\odot}$ to $10^{16}M_{\odot}$ that had previously been stabilised, suddenly find themselves, in a manner of speaking, without any visible means of support. Collapse then proceeds in earnest, and it is the smaller fluctuations, with masses close to $10^{5}M_{\odot}$, that will most rapidly collapse to form virialized systems.

So far, this account of the formation of structure looks promising. Gravitational collapse is essentially arrested until the time of recombination, but soon after this the Jeans mass drops to a value of $10^{5}M_{\odot}$, a mass that is typical for the oldest known stellar systems – the globular clusters. There is however a fundamental problem with this scenario that relates to the level of relative density fluctuation that would be required at recombination in order for this model to produce the structure that we observe in the present-day Universe. Detailed calculations show that this purely baryonic model would require fluctuations in density at recombination that would be detectable now as temperature fluctuations in the cosmic microwave background at the level of about 1 part in 10^3.

- Why is this a problem for this model?
- The observed amplitude of fluctuation in temperature in the cosmic microwave background is about 1 part in 10^5 – about a hundred times too small to give rise to the structures observed in the present-day Universe.

An alternative way of viewing the problem is that there has been insufficient time since recombination for fluctuations to grow from their observed value at that time (1 part in 10^5) to give the structure that we observe today.

We have seen throughout this book that dark matter plays an important dynamical role in many astrophysical systems, so it seems sensible to consider whether dark matter may help in resolving the problem of structure formation.

- If the dark matter was baryonic in nature, could that help in resolving the problem of formation of structure?
- No. We have just seen that any baryonic matter – whether it ends up being luminous or dark – could not give rise to the structure that we see in the present-day Universe.

In the next section we will see how the formation of structure might be modified by the presence of non-baryonic dark matter.

6.6.2 Gravitational collapse with dark matter

In Chapter 2 we saw that numerical simulations of the formation of structure usually incorporate a dark matter component.

It was also noted that hot dark matter scenarios typically cannot reproduce structures across the range of scales that are observed in the present-day Universe. Cold dark matter models seem somewhat more successful, and here we will consider how cold dark matter could play a vital role in forming structure in the Universe.

In the context of cold dark matter, the term 'cold' refers to the fact that the particles have been moving at speeds much slower than the speed of light since very early times in the history of the Universe. This in itself does not help in the formation of structure – since this is essentially the same as the behaviour of the normal baryonic matter in the Universe. As you saw in the previous section, the formation of structure requires a higher degree of density fluctuation at the time of recombination than is observed. This could happen if the cold dark matter particles do not interact to any significant extent with photons or with the baryonic matter. This would mean that the cold dark matter particles would not be supported by the radiation pressure, and so gravitational collapse of density fluctuations of this cold dark matter could start before the time of recombination.

This lack of interaction with photons or electrons is an important characteristic of cold dark matter particles. The term **weakly interacting massive particle** or (**WIMP**) is often used to denote the hypothetical cold dark matter particles. The term 'weakly' refers to the fact that such particles only interact by the weak interaction and do not take part in electromagnetic or strong interactions (they do however respond to gravitational fields). The issue of what WIMPs might actually be is considered in Chapter 8.

In cold dark matter scenarios, primordial fluctuations can start to grow at much earlier times than the epoch of recombination. There is ample time for the density fluctuations in this matter to reach a value of one part in 10^3 at the time of recombination. This level of fluctuation is not evident on the last-scattering surface because there is no significant interaction between WIMPs and photons. Furthermore the behaviour of the baryonic matter is dominated by radiation pressure up until recombination, and thus its distribution appears smooth to one part in 10^5, just as if the Universe contained only baryonic matter.

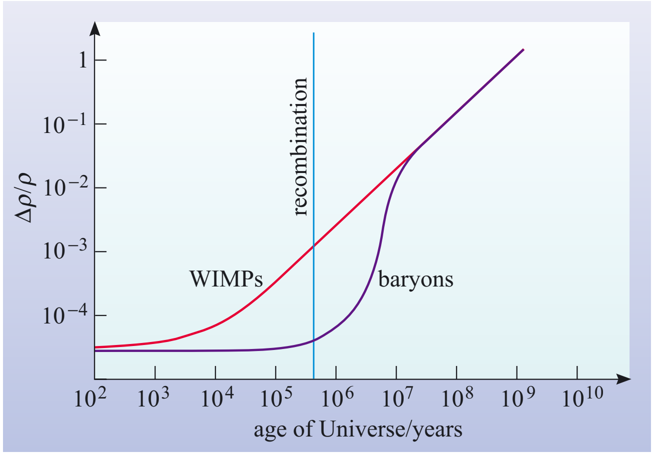
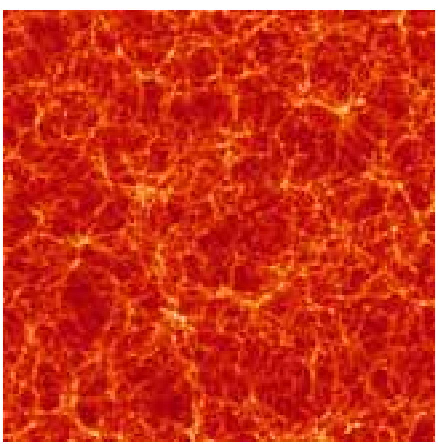


Figure 6.22 The evolution of density perturbations for matter in the form of WIMPs and baryons. The magnitude of a perturbation is characterized by the typical fractional change in density ($\Delta\rho/\rho$). (Adapted from Kolb and Turner, 1990)

After recombination however, the dominant effect on baryonic matter is the gravitational attraction of regions in which over-densities of cold dark matter have accumulated. In a sense the baryonic matter 'falls into' the density enhancements of cold dark matter. This behaviour is summarized in Figure 6.22 which shows the evolution of density enhancements, characterized by the typical relative density fluctuation ($\Delta\rho/\rho$) of matter in the form of WIMPs and in baryons.

Models of the formation of structure under the influence of cold dark matter have undergone scrutiny by conducting computer-based simulations and comparing the resulting structure to that which is measured in the real Universe. The comparison that is made between the outcome of a simulation and real observations usually involves some statistical measure of the structure such as a counts-in-cells analysis as described in Section 4.6. An added, and major, complication is that it is not clear how closely the visible mass in the Universe traces the distribution of dark matter.

Such studies are able to constrain some cosmological parameters relating to the distribution of cold dark matter. For instance, it is possible to rule out the cosmological model which has a mass density parameter $\Omega_m = 1$, no cosmological constant ($\Lambda = 0$), and in which the matter in predominantly in the form of cold dark matter – this is often referred to as the 'standard cold dark matter' (SCDM) model (a simulation of this case is shown in Figure 6.23a). However, other variants of cold dark matter models do produce structure that is similar to that which occurs in


(a)

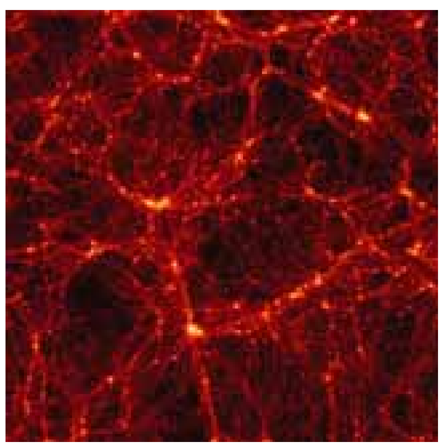

(b)

Figure 6.23 The results of numerical simulations of the formation of structure in two different dark matter scenarios. Both of the models shown have cold dark matter as the dominant contribution to mass. The simulations show the distribution of matter at the present time. (a) A model with no cosmological constant and $\Omega_m = 1$. (b) A model with a non-zero cosmological constant and with $\Omega_m + \Omega_\Lambda = 1$. The model in (a) is incompatible with the observed structure in the real Universe, whereas model (b) gives rise to structure that is similar to that which is observed. (Jenkins *et al.*, 1998/The Virgo Consortium)

the real Universe. For instance one cosmological model with cold dark matter, a non-zero cosmological constant and a flat geometry (the so-called ΛCDM model) appears to be a good representation of reality (Figure 6.23b).

So study of the formation of structure can provide useful information about the nature of dark matter, but it cannot simultaneously determine the values of the cosmological parameters. However, there are other observational techniques that do allow the cosmological parameters to be constrained, and it is to this topic that we turn in the next chapter.

6.7 Summary of Chapter 6

The evolution of the cosmic background radiation

- The cosmic background radiation has a black-body spectrum and its current temperature is about 2.73 K.
- The temperature T of the cosmic background radiation varies with scale factor $R(t)$ according to

$$T \propto \frac{1}{R(t)} \tag{6.6}$$

- At times when the scale factor of the Universe was much smaller than at present, the temperature of the cosmic background radiation would have been much higher.
- At early times the dominant contribution to the energy density of the Universe was that due to radiation. At such times, the temperature is related to time by

$$(T/\mathrm{K}) \approx 1.5 \times 10^{10} \times (t/\mathrm{s})^{-1/2} \tag{6.19}$$

The very early Universe

- Current physical theory breaks down in describing events that took place at, or before, the Planck time ($t \sim 10^{-43}$ s).
- It is speculated that major physical effects could have arisen when grand unification ended ($t \sim 10^{-36}$ s). At this time, the strong and electroweak interactions became distinct. One such effect may be the process of inflation, which resulted in the scale factor increasing very rapidly for a short period of time.
- At early times, the content of the Universe would have been all types of quark and lepton and their antiparticles. There were also particles present that mediate the fundamental interactions (such as the photon), as well as dark-matter particles. There was a slight excess of matter over antimatter.
- At $t \sim 10^{-5}$ s free quarks became bound into hadrons. Most of these hadrons either decayed or annihilated with their antiparticles, leaving only protons and neutrons. For every 10^9 or so annihilation events that occurred, there would have been one proton or neutron left over.
- At $t \approx 0.7$ s neutrinos had their last significant interaction with other particles (apart from the effects of gravity). Shortly after this, electron–positron pairs annihilated, leaving only a residual number of electrons whose summed electric charges exactly balance the charge on the protons.

Primordial nucleosynthesis

- In the first few hundred seconds of the history of the Universe, the physical conditions were such that nuclear fusion reactions could occur. Such reactions led to the formation of deuterium, helium and lithium.
- The first step in the production of helium is the formation of deuterium. This nuclide is unstable to photodisintegration at temperatures above 10^9 K. The formation of helium did not start until $t \approx 225$ s. During this time, some neutrons decayed to protons, and this had an effect on the mass fraction of helium that was produced by primordial nucleosynthesis.
- The mass fraction of helium that is predicted by primordial nucleosynthesis is about 24%. This is in good agreement with measurements of the helium abundance in interstellar gas and stars, and provides very strong evidence to support the hot big bang model.

The scattering of photons and the cosmic microwave background

- The cosmic microwave background that is observed at the present time, is radiation that was last scattered at a redshift of about 1100 ($t \sim 3$ to 4×10^5 years). The radiation appears to originate from the last-scattering surface which is at this redshift.
- The scattering of background radiation photons stopped when the number density of free electrons became very low, and this occurred because of the recombination of electrons and nuclei to form neutral atoms.
- The observed high degree of uniformity of the cosmic microwave background leads to the horizon problem – which is that regions of the last-scattering surface that are more than about 2° apart could not have come into thermal equilibrium by the time that last scattering occurred.
- Our motion relative to co-moving coordinates can be determined by analysis of the observed dipole anisotropy of the cosmic microwave background.
- The cosmic microwave background shows intrinsic anisotropies in temperature at a level of a few parts in 10^5. These anisotropies result from density variations in the early Universe.

The formation of structure

- The formation of structure in the Universe would have proceeded by gravitational collapse from density fluctuations in the early Universe.
- Prior to recombination, the high degree of scattering between photons and electrons prevented density fluctuations in baryonic matter from growing substantially.
- If all matter was baryonic in form, then the level of fluctuation that is observed on the last scattering surface is too small to explain the structure that we observe at the present time.
- The observed level of structure in the present-day Universe can be explained if density fluctuations in non-baryonic matter had begun to grow prior to recombination, and baryons were subsequently drawn into those collapsing clouds of dark matter.

Questions

QUESTION 6.14

Draw a ‘time-line’ for the history of the Universe that indicates the major events that occurred at different times from the Planck time to the present day. Include on this time-line an indication of the temperature at the times of these events.

QUESTION 6.15

Briefly summarize what is meant by a ‘theory of everything’. Why is such a theory required to understand the processes that occurred in the very early Universe?

QUESTION 6.16

During the process of inflation, the scale factor would have increased by an enormous factor. What consequences would this have for the temperature at this time?

QUESTION 6.17

State what the qualitative effect would be on the mass fraction of helium-4 produced by primordial nucleosynthesis if the photodisintegration of deuterium required photons of much higher energy than 2.23 MeV.

QUESTION 6.18

Suppose we received a message from (hypothetical) astronomers in a galaxy that has a current redshift of $z = 2.5$. What would they say they found as the temperature of the cosmic microwave background at the time of their transmission?